



ISSN: 2230-9926

Available online at <http://www.journalijdr.com>

# IJDR

International Journal of Development Research

Vol. 10, Issue, 04, pp. 35416-35421, April, 2020

<https://doi.org/10.37118/ijdr.18717.04.2020>



RESEARCH ARTICLE

OPEN ACCESS

## REDES NEURAIIS ARTIFICIAIS NA PREVISÃO DE CONTÁGIO E ÓBITOS POR COVID-19: UM ESTUDO NO ESTADO DO PARÁ, BRASIL

<sup>1</sup>Jonas Elias Castro da Rocha, <sup>1</sup>Gilberto Nerino de Souza Júnior, <sup>1</sup>Silvana Rossy de Brito, <sup>2</sup>Adriana Ribeiro Carneiro Folador, <sup>2</sup>Rommel Thiago Jucá Ramos, <sup>\*</sup>1Marcus de Barros Braga and <sup>1</sup>Marcel do Nascimento Botelho

<sup>1</sup>Universidade Federal Rural da Amazônia, Pará, Brasil; <sup>2</sup>Universidade Federal do Pará, Pará, Brasil

### ARTICLE INFO

#### Article History:

Received 17<sup>th</sup> January, 2020

Received in revised form

19<sup>th</sup> February, 2020

Accepted 10<sup>th</sup> March, 2020

Published online 30<sup>th</sup> April, 2020

#### Key Words:

Redes Neurais Artificiais, Análise dos Componentes Principais, COVID-19, Saúde Pública, Predição de Série Temporal.

*\*Corresponding author: Jonas Elias Castro da Rocha,*

### ABSTRACT

O coronavírus SARS-CoV-2, anteriormente denominado de 2019-nCoV, é o patógeno causador da Síndrome Respiratória Aguda Grave do Coronavírus 2, amplamente difundida como COVID-19 e possui uma disseminação extremamente rápida, forçando a aplicação de novas políticas públicas para seu enfrentamento. Diversas técnicas de predição de caso futuros de contágio e óbitos foram apresentadas em diferentes regiões do planeta. No presente estudo, propomos um modelo de previsão de curto prazo baseado em Análise dos Componentes Principais e Redes Neurais Artificiais, capaz de estimar o número de casos e de óbitos causados pelo SARS-CoV-2. O modelo adotado é baseado em dados, onde toda inferência é feita a partir do conhecimento descoberto. O estudo apresenta resultados para o estado do Pará e Brasil, com séries temporais que servem de base para analisar o impacto na capacidade de atendimento nos leitos de Unidade de Terapia Intensiva, servindo de suporte à tomada de decisão por parte dos órgãos de vigilância em saúde.

Copyright © 2020, Jonas Elias Castro da Rocha et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Citation: Jonas Elias Castro da Rocha, Gilberto Nerino de Souza Júnior, Silvana Rossy de Brito, Adriana Ribeiro Carneiro Folador et al. "Redes neurais artificiais na previsão de contágio e óbitos por covid-19: um estudo no estado do Pará, Brasil", *International Journal of Development Research*, 10, (04), 35416-35421.

### INTRODUCTION

A doença respiratória COVID-19, causada pelo vírus SARS-CoV-2, veio a desencadear uma pandemia global (Chan *et al.*, 2020), totalizando 2.397.216 casos confirmados e 162.956 óbitos até 21 de abril de 2020, de acordo com a Organização Mundial da Saúde (OMS). No Brasil, já foram contabilizados 43.079 casos confirmados e 2.741 óbitos, segundo o Ministério da Saúde (Saúde, 2020). Os números do estado do Pará estão em 1.026 casos confirmados e 38 óbitos (Pará, 2020). O vírus possui alta capacidade de transmissão e causa sintomas como febre, tosse, fadiga, produção de escarro, falta de ar, dor de garganta e dor de cabeça e, em alguns casos, observam-se sintomas gastrointestinais, como diarreia e vômito. A transmissão de SARS-CoV-2 de humano para humano ocorre principalmente entre membros da família, parentes ou amigos que entrarem em contato com pacientes ou portadores com incubação. Com base nas informações atuais, a maioria dos pacientes apresentou bom prognóstico, enquanto alguns pacientes estavam em estado crítico, especialmente aqueles do

grupo de risco, como idosos e aqueles com doença crônica subjacente (Guo *et al.*, 2020). Várias espécies de coronavírus são considerados patógenos graves para seres humanos, causando doenças respiratórias, hepáticas, gastrointestinais e neurológicas. Estes organismos já foram encontrados em seres humanos, pássaros, gado, ratos, pangolins, morcegos e outros animais selvagens. Os surtos de dois coronavírus anteriores, SARS-CoV e MERS-CoV em 2003 e 2012, respectivamente, comprovaram a transmissão de animal para animal e de humano para humano (Al-qaness *et al.* 2020). Um dos problemas graves da COVID-19 é seu período de incubação de até 14 dias, assim, durante este período o vírus pode ser transmitido (Farida *et al.*, 2020). A disseminação deste patógeno é muito perigosa, o que requer a construção de políticas públicas, além de planos mais rígidos de enfrentamento e contingência (Farida *et al.*, 2020 e Fernandes *et al.*, 2020). Deste modo, torna-se essencial conseguir prever os casos estimados para os dias seguintes, a fim de subsidiar planos de contenção e proteção necessários em tempo hábil, através de medidas como a restrição de viagens,

monitoramento de casos, rastreamento de contatos, quarentena, orientações e informações ao público e desenvolvimento de kits de testes. Ainda existe a necessidade de comprovação da eficácia de medicamentos e desafios quanto ao planejamento da logística de serviços hospitalares, bem como a aquisição de equipamentos de proteção individual e suprimentos hospitalares requeridos para o combate à pandemia (Lu *et al.*, 2020). De acordo com a Secretaria Estado de Saúde do Pará, o primeiro caso de infecção por COVID-19 foi em 18 de março e o primeiro óbito em 01 de abril deste ano. A evolução das infecções tem variado de acordo com as respostas dos governos mediante a crise, em alguns casos parecendo estar contida e em outros atingindo proporções gravíssimas (Guo *et al.*, 2020). Surge então a necessidade de utilização de abordagens orientadas a dados, a partir de informações disponíveis por cidade, por estado, por região, por país, etc., para estimar quantitativamente os casos de contágio e óbitos, além de servir de base para se avaliar as medidas e políticas implementadas a fim de reduzir a propagação e o impacto da doença. As Redes Neurais Artificiais (RNA) podem fazer aproximação de funções contínuas e podem ser usadas com sucesso na previsão de séries temporais com a vantagem de aproximar qualquer função não linear sem nenhuma informação prévia sobre as propriedades da série de dados (Oancea and Cristian, 2013). Métodos clássicos usados para previsão de séries temporais, como Box-Jenkins, ARMA e ARIMA pressupõem que exista uma relação linear entre entradas e saídas (Allende *et al.*, 2002). Como as redes neurais podem aproximar funções não lineares com um conjunto limitado de parâmetros, elas servem como uma ferramenta eficaz para aproximar e estimar os efeitos de uma quarentena em combinação com os modelos epidemiológicos analíticos (Dandekar e Barbastathis, 2020).

O presente trabalho propõe o uso de Redes Neurais Artificiais para predição do número de casos de contágio e óbitos causados por COVID-19, além realizar uma análise da capacidade de suporte com relação aos leitos de Unidade de Terapia Intensiva (UTI) necessários neste período de pandemia. O estudo foi realizado no estado do Pará, região norte do Brasil, a partir de dados públicos do Ministério da Saúde. A modelagem deste fenômeno epidemiológico não é exclusiva do estado do Pará, então, partindo do comportamento da doença no mundo, nos territórios continentais, no Brasil, nas regiões do Brasil, nos estados e nos municípios, foi possível realizar predições para o estado do Pará em uma abordagem orientada a dados. Logo, a RNA treinada a partir dos dados coletados pode inferir e prever novos casos epidemiológicos, podendo servir de suporte à tomada de decisão por parte dos órgãos de vigilância em saúde e autoridades competentes.

## MATERIAIS E MÉTODOS

**Base de Dados:** Os dados utilizados neste estudo são públicos, fornecidos pela Secretaria de Estado de Saúde do Pará (<http://www.saude.pa.gov.br/coronavirus/>), pelo Ministério da Saúde (<https://saude.gov.br/>) e pelos relatórios da Organização Mundial da Saúde (<https://www.who.int/docs/>). Foram identificados os padrões comportamentais da sociedade, extraídos para análise os casos ocorridos no estado do Pará e comparados com os estratos nacionais e mundiais. Com base no banco de dados fornecido pelo órgão estadual de saúde, foi usada a série histórica com informações diárias das

ocorrências a partir de 18/03/2020, data do primeiro caso de COVID-19 confirmado no Pará, até o dia 21/04/2020.

**Análise de Componentes Principais:** Para se determinar as variáveis que seriam utilizadas na composição dos neurônios artificiais da camada de entrada das redes neurais, foi aplicado o método de Análise de Componentes Principais (ACP), que processou o conjunto de informações exemplificadas na Tabela 1, onde 17.237 linhas e 09 colunas (variáveis) totalizam 153.135 observações. As variáveis foram:

1. Tipo de Estratos: city, country, region, state, territory, world;
2. Territórios Continentais: South Region of the America, European Region, Regions of Americas, Western Pacific Region, South-East Asian Region;
3. País: Afeganistão, África do Sul, .... Brasil .... Zâmbia, Zimbábue;
4. Regiões do Brasil: CENTRO-OESTE, NORTE, NORDESTE, SUDESTE, SUL;
5. Estado: AC, AL, AP, AM, BA, CE, ... , SP, SE, TO;
6. Cidade: Abaeté, Abaetetuba, Abaiba, Abreu e Lima, Açailândia, Acará, ... , Zacarias, Zé Doca, Zortéa;
7. Dias Após o Início da Contaminação: numérico inteiro;
8. Casos Confirmados: numérico inteiro;
9. Óbitos: numérico inteiro;

Para tanto, os dados foram padronizados para identificação das variáveis latentes ortogonais, que representam a variação máxima dos dados originais que não estão correlacionados entre si. As variáveis latentes foram determinadas a partir da combinação linear entre os valores padronizados das variáveis utilizadas na ACP e seus respectivos autovetores de cada dimensão multivariada (Equação 1).

$$\begin{aligned} Y_1 &= e_{11} \frac{X_1 - \mu_1}{\sigma_1} + e_{12} \frac{X_2 - \mu_2}{\sigma_2} + \dots + e_{1p} \frac{X_p - \mu_p}{\sigma_p} \\ Y_2 &= e_{21} \frac{X_1 - \mu_1}{\sigma_1} + e_{22} \frac{X_2 - \mu_2}{\sigma_2} + \dots + e_{2p} \frac{X_p - \mu_p}{\sigma_p} \\ Y_p &= e_{p1} \frac{X_1 - \mu_1}{\sigma_1} + e_{p2} \frac{X_2 - \mu_2}{\sigma_2} + \dots + e_{pp} \frac{X_p - \mu_p}{\sigma_p} \end{aligned} \quad (1)$$

onde,  $Y$  = variável latente,  $X$  = vetor aleatório da variável latente,  $\mu$  = valor médio da variável latente,  $\sigma$  = valor do desvio padrão da variável latente e  $e$  = autovalores  $p$  originados. Conforme proposto por (Mingoti, 2005), os componentes principais selecionados foram os primeiros cuja soma dos autovetores foi capaz de explicar o percentual mínimo de 70% da variação acumulada.

**Arquitetura, Treinamento e Validação das RNAs:** As variáveis consideradas para o treinamento e validação das RNAs, considerando-se as entradas e saídas dos ciclos de treinamento, são qualitativas, quantitativas e discretas. As variáveis que compõem os neurônios da camada de saída das RNAs são os Casos Confirmados (CC) e os Óbitos (OB) por COVID-19. O intervalo do número de neurônios na camada oculta foi definido pelo método de Fletcher-Gloss (Silva *et al.*, 2010), mostrado na Equação 2.

$$n = 2 \times \sqrt{n} + n_2 \leq n_1 \leq 2 \times n + 1 \quad (2)$$

onde  $n$  = número de entradas da rede,  $n_1$  = quantidade de neurônios na camada oculta e  $n_2$  = quantidade de neurônios na camada de saída.

Tabela 1. Conjunto de Variáveis Testadas na ACP para Fazer Parte da Camada de Entrada das RNAs

N	Tipo de Estrato	Territórios Continentais	País	Regiões do Brasil	Estado	Cidade	Dias Após o Início da Contaminação	Casos Confirmados	Óbitos
1	city	South Region of the America	Brasil	SUDESTE	MG	Abaeté	0	CC	OB
2				NORTE	PA	Abaetetuba	1		
3				NORDESTE	CE	Abaíra	2		
4				NORDESTE	PE	Abreu e Lima	3		
5				NORDESTE	MA	Açailândia	4		
6				NORTE	PA	Acará	5		
7				NORDESTE	RN	Acari	6		
8				NORDESTE	CE	Acopiara	7		
9				NORTE	AC	Acrelândia	8		
10				SUDESTE	MG	Açucena	9		
11				SUDESTE	SP	Adamantina	10		
12				NORDESTE	BA	Adustina	11		
13				SUDESTE	ES	Afonso Cláudio	12		
14				SUDESTE	SP	Águas de Lindóia	13		
15				SUDESTE	SP	Águas de São	14		
16				CENTRO-	GO	Águas Lindas de	15		
17				SUL	SC	Águas Mornas	16		
18				SUDESTE	SP	Agudos	17		
19				NORDESTE	BA	Aiquara	18		
20				SUL	RS	Ajuricaba	19		
21				NORDESTE	BA	Alagoinhas	20		
22				SUDESTE	SP	Alambari	21		
23				CENTRO-	MS	Alcinópolis	22		
24				SUL	RS	Alegrete	23		
25				SUDESTE	MG	Alfenas	24		
26				SUDESTE	ES	Alfredo Chaves	25		
27				NORDESTE	PE	Aliança	26		
...				...	...	...	27		
...				...	...	...	28		
...				...	...	...	29		
172				NORDESTE	MA	Zé Doca	30		

A partir deste ponto, o próximo passo foi definir a arquitetura ideal da Rede Neural Artificial para o conjunto de dados obtido e, para tal, vários testes foram realizados. Inicialmente, dois tipos de RNA foram treinados, uma rede *Multilayer Perceptron* (MLP) e uma rede *Radial Basis Function* (RBF) (Haykin, 1994). De forma geral, tanto MLPs quanto RBFs possuem arquiteturas que consistem em uma camada de entrada, que recebe as variáveis preditoras, uma ou mais camadas ocultas, que executam o processamento paralelo, e uma camada de saída, que contém a(s) variável(is) predita(s). No presente estudo, a variável DAN (composta por TC = Território Continental, CC = Casos Confirmados, OB = Óbitos, P = País, UF = Unidade Federativa, RB = Regiões do Brasil, CB = Cidades do Brasil e EV = Estrato da Variável) é processada nos neurônios da camada de entrada das RNAs e as variáveis CC e OB nos neurônios da camada de saída. As funções de ativação experimentadas na camada oculta foram: Gaussiana, Tangente Hiperbólica, Logística, Exponencial e Senoidal. Na camada de saída foram testadas as funções Identidade, Tangente Hiperbólica, Logística e Senoidal. Os algoritmos de treinamento usados foram: Broyden-Fletcher-Goldfarb-Shanno (BFGS) para o MLP e Redundant Byzantine Fault Tolerance (RBFT) para a RBF. Utilizou-se também a função SOS (Sum of Squares) para correção dos erros e uma taxa de decaimento dos pesos de 0,0001 na camada oculta. Foram treinadas vinte RNAs, das quais se selecionou a rede com melhores medidas de precisão do ajustamento, com base no Coeficiente de Correlação Linear de Pearson entre os

valores das variáveis reais com os valores das preditas, de treinamento e validação; erro de treinamento e validação; Análise de Sensibilidade Global, Raiz Quadrada do Erro Médio Normalizado em % (NRMSE) e; Dispersão de Resíduos. As RNAs foram construídas com a ferramenta Automatic Network Search do software Statistica, versão 13.2 Trial (StatSoft I., 2004). Para validação das previsões de CC e OB pela RNA selecionada, aplicou-se o teste qui-quadrado ( $\chi^2$ ), com significância estatística de 5%. O conjunto de observações utilizado para validar a modelagem da RNA não fez parte da fase de treinamento e ajuste da rede, assim, os vieses nas previsões foram evitados.

**Análise da Capacidade de Suporte:** Para analisar a capacidade de suporte do Estado do Pará com relação aos leitos de Unidade de Terapia Intensiva (UTI) necessários para o tratamento de infectados com SARS-COV-2 sob condições de Síndrome Respiratória Aguda Grave (SRAG), utilizou-se a taxa de uso de leito de UTI confrontando-a com os valores preditos pelas RNAs. Este banco de dados, bem como a quantidade de leitos de UTI para uso no combate a COVID-19 é oriundo de uma série de dados históricos semanais disponibilizados pelo Ministério da Saúde (<https://covid.saude.gov.br/>).

**Erro Geral:** Em modelagem estatística, uma maneira comum de se medir a qualidade do ajuste do modelo é através do RMSE (Root Mean Square Error) ou Desvio Quadrático

Médio da Raiz. Se as respostas previstas estiverem muito próximas das respostas verdadeiras, o RMSE terá um valor próximo de zero. Neste estudo, o NRMSE (Normalized Root Mean Square Error) foi usado para medir a acurácia geral do modelo neural, cujo o uso remete ao RMSE normalizado entre 0 e 1, expressando a porcentagem de erro do modelo preditor em relação ao conjunto de dados observados real. O RMSE e NRMSE são apresentados na Equação 3 e Equação 4:

$$RMSE = \sqrt{\sum_{i=1}^N \frac{(P_i - R_i)^2}{N}} \quad (3)$$

$$NRMSE = \frac{RMSE}{R_{max} - R_{min}} \quad (4)$$

onde,  $N$  é o número total de observações,  $R_i$  é o valor para  $i$ -ésima observação real,  $P_i$  é a previsão para a  $i$ -ésima observação do modelo preditor utilizado,  $R_{max}$  é o valor máximo do vetor  $R$ , e  $R_{min}$  é o valor mínimo do vetor  $R$ .

aos componentes principais (CPs) gerados e a porcentagem de variância acumulada. No caso, o primeiro e o segundo componentes principais são responsáveis por 42,97% e 27,64% da variação dos dados, respectivamente, totalizando juntos, 70,62% da variação encontrada. A ACP identificou dois grupos de variáveis. Assim, buscando-se entender as máximas correlações entre as variáveis no comportamento da curva de predição de dados para o estado do Pará, procedeu-se com o treinamento e validação das RNAs em duas etapas. As variáveis agrupadas em CP1, quais sejam, TC, CC, OB, EV e P determinaram os pesos sinápticos na etapa de processamento dos neurônios da camada de entrada no primeiro grupo de RNAs treinadas. Já, as agrupadas em CP2, variáveis RB, UF e CB, foram processadas nos neurônios da camada de entrada do segundo grupo de RNAs (Tabela 3). Os parâmetros estatísticos das redes neurais de melhor performance para os dados são apresentadas na Tabela 3, na qual, verifica-se que, para o conjunto de variáveis mundiais, o melhor modelo foi o MLP 8-12-2, com 8 (oito) neurônios de entrada, 12 (doze) na camada

**Tabela 2. Resultado da Análise de Componentes Principais (ACP) para as Variáveis de Entrada**

Componentes Principais	Autovetores	Variância Total	Variância Acumulada	Variáveis	Score
		%	%		
1	3,4378	42,9729	42,9729	TC	-0,7616
				CC	-0,7517
				OB	-0,7211
				EV	-0,9317
				P	-0,8644
2	2,2115	27,6433	70,6162	RB	-0,7317
				UF	-0,8955
				CB	-0,9184

**Tabela 3. Resultado da Análise de Componentes Principais (ACP) para as Variáveis de Entrada**

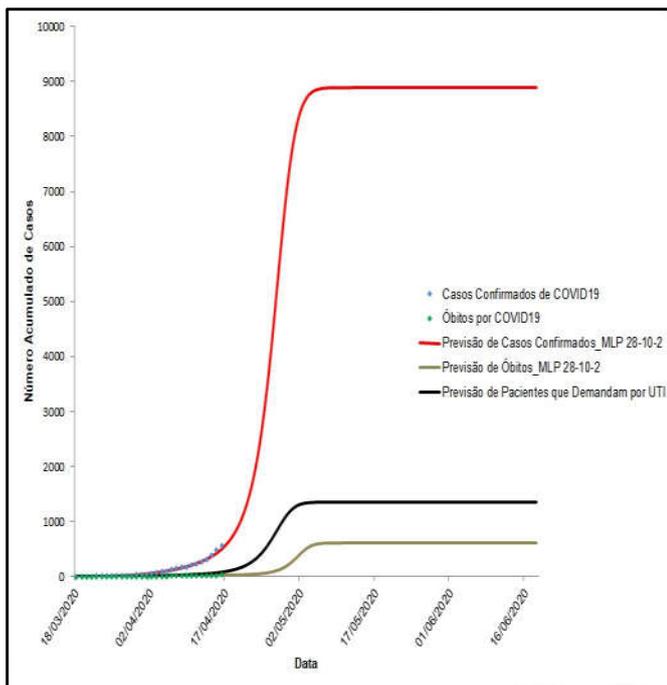
Características das RNAs Seleccionadas		Estado do Pará	Mundo
Arquitetura		MLP 28-10-2	MLP 8-12-2
Treinamento	Correlação	0,99866	0,98064
	Erro	0,00002	0,98185
Validação	Correlação	0,99702	0,00150
	Erro	0,00007	0,00175
Algoritmo de Treinamento		BFGS 10000	BFGS 3927
Função de Ativação	Camada Oculta	Exponential	Logistic
	Camada de Saída	Logistic	Identity
Teste Qui-Quadrado $p$ - value	Casos Confirmados	0,20453	0,24710
	Óbitos	0,99998	0,14560
NRMSE	Casos Confirmados	6,30157%	0,77781%
	Óbitos	1,06001%	0,00014%
Análise de Sensibilidade Global	Casos Confirmados	139,22261	938,65742
	Óbitos		

## RESULTADOS

Com o objetivo de identificar os principais atributos a serem utilizados na RNA, foi executado o módulo de Análise de Componentes Principais. Os autovetores, indicam que a expressão das variáveis no primeiro componente principal foi maior para "Tipo de estratos", "Territórios Continentais" e "País". Já, no segundo componente principal, as variáveis que apresentaram maior expressão foram as "Regiões do Brasil", "Estado" e "Cidade". Desta maneira, foram definidos os dois grupos de variáveis a serem incluídas nas RNAs em processamentos paralelos, um grupo de variáveis mundiais e outro nacional. A Tabela 2 mostra os valores obtidos na ACP para os autovetores, a porcentagem de variância total associada

oculta e 2 (dois) na camada de saída, de maneira que, o ajuste desse primeiro grupo de variáveis auxiliou na modelagem do fenômeno com múltiplos neurônios em processamento e, portanto, uma melhor estimativa da curva extraída para análise dos dados do estado do Pará. No segundo grupo de variáveis, o modelo de melhor performance para a estimativa de CC e OB foi o da arquitetura MLP 28-10-2, com 28 (vinte e oito) neurônios de entrada, 10 (dez) na camada oculta e 2 (dois) na camada de saída. A rede foi treinada com algoritmo BFGS, função de correção de erro SOS, função de ativação Exponencial na camada oculta e função de ativação Logística na camada de saída. Na fase de validação, todas as RNAs seleccionadas na fase de treinamento apresentaram, no teste Qui-Quadrado, o valor probabilísticos não significativo ao

nível de significância de 0,05, assim, é possível dizer que o conjunto de dados reais é estatisticamente igual ao conjunto de dados estimados pelas RNAs, tanto para a quantidade de casos confirmados de COVID-19 quanto para o número de óbitos, nas avaliações ao nível mundial com aprendizado comportamental para as RNAs treinadas ao conjunto de dados extraído para o Estado do Pará. Para os dados do estado do Pará, a Rede Neural apresentou um erro geral (NRMSE) de 6,30157% para os Casos Confirmados (CC) e 1,06001% para os Óbitos (OB). Já para os dados mundiais, a Rede Neural apresentou um erro geral (NRMSE) de 0,77781% para os Casos Confirmados (CC) e 0,00014% para os Óbitos (OB).



**Figura 1. Série Temporal diária referente ao número acumulado de Casos Confirmados (marcadores azuis) e de Óbitos (marcadores verdes), sobrepostos pelas suas respectivas curvas de previsão ajustadas pela RNA, em linhas de cor vermelha e marrom. Foram adicionados também a predição por número de pacientes que demandam UTIs no Estado do Pará**

A Figura 1 apresenta a série temporal gerada pelas RNAs. Ao se verificar a ascendência da curva no seu pleno crescimento, estima-se que o período máximo de contágio por COVID-19 no estado do Pará ocorra entre os dias 26 a 28 de abril. Para a curva ajustada de casos confirmados, o ponto máximo culmina em 8.895 casos entre os dias 09 a 12 de maio. A partir de então, verificou-se uma estabilização da curva e redução de novos casos. Para o número acumulado de óbitos, o ponto máximo previsto ocorrerá entre os dias 03 a 05 de maio do ano corrente, de 500 a 600 óbitos no estado. O modelo proposto, que faz previsão de curto prazo (Short-Term Forecasting), tem capacidade de prever o número de casos confirmados e óbitos em até 30 dias após a notificação do último caso de COVID-19. Considerando-se que o ajuste da curva ao banco de dados treinados seguirá o modelo teórico da rede selecionada, o comportamento da sociedade definirá a ascendência ou a estagnação do fenômeno modelado. Conforme dados informados pela Secretaria de Estado de Saúde do Pará, a demanda de pacientes com síndrome respiratória grave internados em leitos de unidade de terapia intensiva (UTI) com COVID-19 é de 16,39% do total dos casos diagnosticados com a doença. Desta maneira, as

extrapolações não supervisionadas das RNAs estimam que, aproximadamente 1.360 leitos de UTIs sejam necessários para suprir as demandas por esse serviço no estado.

## DISCUSSÃO

Muitos estudos já foram apresentados para prever comportamentos em epidemias. Modelos que efetuam previsão de séries temporais utilizando MLP (Multilayer Perceptron), TDNN (Time-Delay Neural Networks), RBF (Radial Basis Function) e ARIMA (Auto-Regressive Integrated Moving) são adotados com frequência (Ture and Kurt, 2006). Além destes, modelos de previsão baseados em filtros Kalman já foram usados em surtos sazonais de influenza (Shaman and Karspeck, 2012). Modelos dinâmicos de inferência Bayesiana foram usados para prever surtos de Ebola em vários países africanos, como Libéria, Serra Leoa e Guiné (Shaman, Yang and Kandula, 2014). Redes Neurais Artificiais frequentemente auxiliam a identificar diversas patologias em pacientes (Kuretzki *et al.* 2019). O ANFIS (Adaptive Neuro-Fuzzy Inference System) já foi amplamente aplicado em problemas de previsão e previsão de séries temporais e mostrou bom desempenho em muitas aplicações existentes na determinação da não linearidade nos dados de séries temporais no surto da SARS (Massad *et al.*, 2005) e mais recentemente ao surto de COVID-19 em Wuhan, na China (Al-qaness *et al.*, 2020). Na literatura, as RNAs vêm sendo utilizadas com sucesso como metodologias preditoras em epidemias, em especial na análise de período de quarenta por COVID-19 (Dandekar e Barbastathis, 2020).

Bullock *et al.* (2020) indica que técnicas de inteligência artificial, em especial as Redes Neurais Artificiais, podem ser promissoras em vários aspectos do conhecimento para auxiliar no combate a pandemia do COVID-19. A partir da análise dos resultados, constatou-se que a Rede Neural Artificial selecionada (MLP 28-10-2), ajustada para as condições do estudo (treinada com os dados de todas as unidades da federação e aplicada para prever os casos no estado do Pará) mostrou-se eficaz na previsão de novos casos de COVID-19, bem como no número de óbitos no estado do Pará. O erro geral do modelo proposto (NRMSE) foi considerado baixo, sendo de 6,30157% para os casos confirmados e 1,06001% para os óbitos, para o estado do Pará. Desta maneira, a distribuição percentual do erro de maneira padronizada ficou próxima de zero, considerando que a padronização do erro foi realizada com vista na comparação de arquiteturas de redes que utilizam variáveis diferentes em suas arquiteturas treinadas. Este erro estimado indica um valor satisfatório para o modelo proposto.

Com relação ao número de leitos de UTIs, a rede neural prevê a necessidade de mais de mil leitos de UTIs necessários para suprir as demandas no estado. É um número alto para qualquer cidade do mundo. Todavia, o modelo proposto reflete um cenário comportamental da sociedade, informado por dados oficiais do estado, seja ele com subnotificações ou não, essas informações são, de maneira geral, coletadas de pacientes que foram motivados a fazer o teste por conta do agravamento no quadro sintomatológico da doença. Cabe ressaltar que o modelo adotado é baseado em dados (*Data Driven Modeling*), isto é, toda inferência é feita a partir do conhecimento prévio descoberto nos dados. Assim sendo, os casos não notificados ainda, em sua grande maioria, podem vir a alterar a taxa de letalidade geral da doença. Não pode se deixar de levar em conta que a adesão, maior ou menor, da população em geral ao

distanciamento social preconizado pelas autoridades sanitárias, pode vir a ser determinante para um alargamento do contágio e, por conseguinte, da letalidade do vírus. Ademais, os números apresentados nesta pesquisa corroboram com as expectativas do governo federal e com outros estudos, quanto à estimativa de pico de contaminação, bem como ao período necessário para a quarentena, esperando que a mesma finalize entre meados do mês de maio e início do mês de junho. A modelagem de fenômenos epidemiológicos é sensível ao erro quando feitas estimativas para períodos muito esparsos, por isso foi adotado um modelo preditor de curto prazo. Assim, para se estimar em períodos mais longos, novos treinamentos devem ser realizados para maior acurácia do modelo e, desta forma ser efetivo no apoio à tomada de decisões estratégicas no combate a pandemia por COVID-19.

### Agradecimentos

À Universidade Federal Rural da Amazônia (UFRA) pela oportunidade e idealização deste estudo. À Pró-Reitoria de Pesquisa e Desenvolvimento Tecnológico (PROPED) da UFRA pela ajuda no financiamento deste estudo. Ao Governo do Pará e à SESPAs pelas informações disponibilizadas nos canais oficiais e pelas efetivas medidas de prevenção ao novo Coronavírus.

### REFERÊNCIAS

- Allende, H., Moraga, C., and Salas, R. (2002). Artificial Neural Networks in Time Series Forecasting: A Comparative Analysis. *KYBERNETIKA - Volume 88* (2002), Number 6, Pages 685 - 707.
- Al-qaness, M. A. A., Ewees, A. A., Fan, H., & Abd El Aziz, M. (2020). Optimization Method for Forecasting Confirmed Cases of COVID-19 in China. *Journal of Clinical Medicine*, 9(3), 674. <https://doi.org/10.3390/jcm9030674>
- Bullock, J., Alexandra, Luccioni, Pham, K. H., Lam, C. S. N., & Luengo-Oroz, M. (2020). Mapping the Landscape of Artificial Intelligence Applications against COVID-19. Eprint ArXiv:2003.11336. Retrieved from <http://arxiv.org/abs/2003.11336>
- Chan, J. F. W., Yuan, S., Kok, K. H., To, K. K. W., Chu, H., Yang, J., ... Yuen, K. Y. (2020). A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *The Lancet*, 395(10223), 514–523. [https://doi.org/10.1016/S0140-6736\(20\)30154-9](https://doi.org/10.1016/S0140-6736(20)30154-9)
- Dandekar, R., & Barbastathis, G. (2020). Neural Network aided quarantine control model estimation of global Covid-19 spread. Eprint ArXiv:2004.02752. Retrieved from <http://arxiv.org/abs/2004.02752>
- Fariba Rezaeetalab, Mahnaz Mozdourian, Mahnaz Amini, Zahra Javidarabshahi, F. A. (2020). COVID-19: A New Virus as a Potential Rapidly Spreading in the Worldwide. *Journal of Cardio-Thoracic Medicine*, 8(1), 563–564.
- Fernandes, E. G., Da, J., Santos, S., Keico, H., & Iii, S. (2020). Investigação de surto em navio de carga em tempo de COVID-19, Porto de Santos, Brasil. *Rev. Saúde Pública*, 54. <https://doi.org/10.11606/s1518-8787.2020054002461>
- Guo, Y-R., Cao, Q-D., Hong, Z-S., Tan, Y-Y., Chen, S-D., Jin, H-J., Tan, K-S., Wang, D-Y. and Yan, Y. (2020). The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak – an update on the status, *Military Medical Research* (2020) 7:11 <https://doi.org/10.1186/s40779-020-00240-0>
- Haykin, S. S. (1994). *Neural networks: a comprehensive foundation* (1st ed.). NY: Macmillan College Publishing.
- Kuretzki, C. H., Simão, J., Pinto, P., Beltrão, C. J., & Hamerschmidt, R. (2019). APPLICATION OF NEURAL NETS FOR THE IDENTIFICATION OF OTORHINOLARYNGOLOGICAL DISEASES. *International Journal of Development Research*, 09(11), 32002–32005. Retrieved from <http://www.journalijdr.com>
- Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., et al. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet*, 395(10224), 565–574. [https://doi.org/10.1016/S0140-6736\(20\)30251-8](https://doi.org/10.1016/S0140-6736(20)30251-8)
- Massad, E.; Burattini, M.N.; Lopez, L.F.; Coutinho, F.A. (2005). Forecasting versus projection models in epidemiology: The case of the SARS epidemics. *Med. Hypotheses* 2005, 65, 17–22.
- Mingoti, S. A. (2005). Análise de dados através de métodos de estatística multivariada uma abordagem aplicada. UFMG: Editora UFMG.
- Oancea, B., & Cristian Ciucu, Ş. (2013). Time Series Forecasting Using Neural Networks. *Challenges of the Knowledge Society. IT in Social Sciences*, 1402–1408.
- Pará, S. de E. da S. do. (2020). Boletim Informativo: Coronavírus no Pará. Retrieved from <http://www.saude.pa.gov.br/coronavirus/>
- Saúde, M. da. (2020). Painel de casos de doença pelo coronavírus 2019 (COVID-19) no Brasil pelo Ministério da Saúde. Retrieved from <https://www.saude.gov.br/images/pdf/2020/fevereiro/21/2020-02-21-Boletim-Epidemiologico03.pdf>
- Shaman, J.; Karspeck, A. (2012). Forecasting seasonal outbreaks of influenza. *Proc. Natl. Acad. Sci. USA* 2012, 109, 20425–20430.
- Shaman, J.; Yang, W.; Kandula, S. (2014). Inference and forecast of the current West African Ebola outbreak in Guinea, Sierra Leone and Liberia. *PLoS Curr.* 2014, 6.
- Silva, I. N., Spati, D. H., & Flauzino, R. A. (2010). Redes Neurais Artificiais para Engenharia e Ciências Aplicadas. Curso Prático. Artliber. São Paulo.
- STATSOFT INC. (2004). *Statistica*: data analysis software system, version 7. Tulsa. Disponível em <http://www.statsoft.com/>
- Ture, M.; Kurt, I. (2006). Comparison of four different time series methods to forecast hepatitis A virus infection. *Expert Syst. Appl.* 2006, 31, 41–46.

\*\*\*\*\*