

**UNIVERSIDADE NOVE DE JULHO – UNINOVE**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA E GESTÃO DO**  
**CONHECIMENTO**

**DIEGO OLIVA**

**MODELAGEM AUTORREGRESSIVA NÃO LINEAR DO IMPACTO DO**  
**DISTANCIAMENTO SOCIAL NOS CASOS DE COVID-19 NO ESTADO DE SÃO**  
**PAULO, BRASIL**

São Paulo  
2023

**DIEGO OLIVA**

**MODELAGEM AUTORREGRESSIVA NÃO LINEAR DO IMPACTO DO  
DISTANCIAMENTO SOCIAL NOS CASOS DE COVID-19 NO ESTADO DE SÃO  
PAULO, BRASIL**

Dissertação apresentada ao Programa de Pós-Graduação em Informática e Gestão do Conhecimento da Universidade Nove de Julho – PPGI/UNINOVE - como requisito parcial para obtenção do título de mestre.

Orientador: Prof. Dr. Fabio Henrique Pereira

São Paulo  
2023

Oliva, Diego.

Modelagem autorregressiva não linear do impacto do distanciamento social nos casos de Covid-19 no estado de São Paulo, Brasil. / Diego Oliva. 2023.

74 f.

Dissertação (Mestrado) - Universidade Nove de Julho - UNINOVE, São Paulo, 2023.

Orientador (a): Prof. Dr. Fábio Henrique Pereira.

1. Covid-19. 2. NARX. 3. Modelos autorregressivos. 4. Distanciamento social. 5. Redes neurais.

I. Pereira, Fábio Henrique. II. Título.

CDU 004



Dedico este trabalho ao Dr. Dimas Tadeu  
Covas, que não poupou esforços na busca  
por vacinas durante a pandemia de COVID-  
19

## **AGRADECIMENTOS**

A conclusão deste trabalho representa o término de mais um ciclo de aprendizado, o qual seria impossível sem a ajuda do estimado orientador Prof. Dr. Fábio Henrique Pereira, que se fez presente durante este percurso.

Agradeço aos amigos que me apoiaram, me ajudaram e acreditaram em mim durante este período. O mestrado é uma fase especial na carreira de pesquisadores, pois o mesmo nos apresenta a intensa caminhada do saber, onde um cada pequeno passo na ciência é importante, pois este, somado a outros, leva a humanidade a um mundo melhor.

Aos professores que me acompanharam durante estes anos, pois compartilharam conhecimentos importantes que diferenciaram minha vida.

À diretoria do Programa de Pós Graduação em Informática e Gestão do Conhecimento pela forma como conduzem este trabalho, incentivando os alunos a sempre buscarem o melhor.

Meu muito obrigado.

## RESUMO

Os primeiros casos da doença do coronavírus SARS-CoV-2 (*coronavirus disease - COVID-19*) foram detectados na China em dezembro de 2019, podendo ocasionar uma síndrome respiratória aguda severa (*SARS - Severe Acute Respiratory Syndrome*). O desconhecimento de tratamento preventivo e, inicialmente, a não existência de vacinas, obrigaram as autoridades a adotarem políticas rígidas de isolamento e distanciamento social como, por exemplo, o fechamento de escolas e restrições de uso de espaços públicos. Tais medidas tiveram o objetivo de retardar/evitar a propagação da doença, mas possuíam potencial de gerar impactos econômicos, políticos e culturais. Consequentemente, a plena compreensão dos impactos dessas medidas na propagação da doença ainda demanda investigação. Assim, este trabalho propõe uma modelagem baseada numa rede neural autorregressiva não linear (NARX, do inglês *nonlinear autorregressive with exogenous inputs*) que visa relacionar o distanciamento social ao número de novos casos e óbitos por COVID-19, com base em dados reais de São Paulo. O modelo é criado com dados de antenas de celular que indicam deslocamento dos aparelhos móveis, obtidos pelo Sistema de Monitoramento Inteligente de São Paulo. Foram realizados experimentos visando a calibragem do modelo, por meio da variação do número de camadas e neurônios por camada, do algoritmo de treinamento e do atraso na série temporal a ser considerado no modelo autorregressivo. Posteriormente, foram definidos cenários hipotéticos de distanciamento social para avaliar a influência dessa variável no número de novos casos e óbitos pela doença, com resultados que reforçam a importância dessas medidas de contenção. Os melhores resultados foram obtidos a partir dos dados padronizados com  $MSE=1,009e-06$  para previsão dos novos casos.

Palavras-chave: COVID-19, NARX, Modelos autorregressivos, Distanciamento Social, Redes Neurais

## **ABSTRACT**

The first cases of the SARS-CoV-2 coronavirus disease (COVID-19) were detected in China in December 2019, potentially causing severe acute respiratory syndrome (SARS). The lack of preventive treatment and initially the absence of vaccines forced authorities to adopt strict policies of isolation and social distancing, such as school closures and restrictions on the use of public spaces. These measures aimed to slow down or prevent the spread of the disease but had the potential to generate economic, political, and cultural impacts. Consequently, the full understanding of the impacts of these measures on the spread of the disease still requires investigation. Therefore, this study proposes modeling based on a nonlinear autoregressive neural network with exogenous inputs (NARX) that aims to relate social distancing to the number of new cases and deaths from COVID-19, based on real data from São Paulo. The model is created using data from cell phone antennas that indicate the movement of mobile devices, obtained from the Intelligent Monitoring System of São Paulo. Experiments were conducted to calibrate the model by varying the number of layers and neurons per layer, the training algorithm, and the delay in the time series to be considered in the autoregressive model. Subsequently, hypothetical scenarios of social distancing were defined to evaluate the influence of this variable on the number of new cases and deaths from the disease, with results that reinforce the importance of these containment measures. The best results were obtained from standardized data with  $MSE=1.009e-06$  to predict new cases.

Keywords: COVID-19, NARX, Autoregressive Models, Social Distancing, Neural Networks



## LISTA DE ABREVIATURAS E SIGLAS

COVID-19	Coronavirus disease – 2019
UTI	Unidade de Terapia Intensiva
AR	Modelo Autorregressivo
ARIMA	Modelo Autorregressivo de médias móveis
SARIMA	Modelo Autorregressivo de médias móveis com sazonalidade
NARX	Non-linear autorregressive with exogenous inputs
SIR	Suscetíveis-Infectados-Recuperados
LSTM	Long-short term memory
RT-PCR	Real-time protein c Reative
AUC	Área under the ROC curve
IDEA	incidence decay with exponential adjustment
DS	Distanciamento Social
SARS-COV2	Síndrome Respiratória Aguda Grave coronavírus 2
RMSE	Root Mean Squared Error
MAPE	Mean Absolute Percentage Error
MAE	Mean Absolute Error
MSE	Mean Squared Error
$R^2$	Coeficiente de determinação
SIMI-SP	Sistema de Monitoramento Inteligente de São Paulo
MSE	Erro quadrático médio
IoT	Internet das coisas
RNA	Rede Neural Artificial
IACOV-BR	Inteligência Artificial para COVID-19 no Brasil
WHO	World Health Organization

## LISTA DE FIGURAS

FIGURA 1 Rede neural artificial NARX .....	26
FIGURA 2 Etapas dos procedimentos experimentais .....	28
FIGURA 3 Exemplo de dados obtidos do Google COVID-19.....	30
FIGURA 4 Novos casos de COVID-19 e Média Móvel dos últimos 7 dias .....	32
FIGURA 5 Número de óbitos por COVID-19 e média móvel dos últimos 7 dias .....	32
FIGURA 6 Índice de distanciamento social e média móvel dos últimos 7 dias .....	33
FIGURA 7 Exemplo da arquitetura do modelo NARX .....	34
FIGURA 8 Esquema de treinamento da rede neural.....	35
FIGURA 9 Esquema de rede neural para os cenários de distanciamento social .....	41
FIGURA 10 Comparação entre valores alvo e preditos com erros de previsão .....	44
FIGURA 11 Histograma dos erros de previsão dos modelos .....	45
FIGURA 12 Resultados de previsão com malha aberta hor. de previsão N=7.....	47
FIGURA 13 Resultados de previsão com malha aberta hor. de previsão N=14.....	48
FIGURA 14 Resultados de previsão com malha aberta N=7 .....	49
FIGURA 15 Resultados de previsão com malha aberta N=14 .....	50
FIGURA 16 Resultados de previsão com malha fechada N=14.....	52
FIGURA 17 Resultados de previsão N=14. Modelos usados em malha fechada .....	53
FIGURA 18 Previsão múltiplos passos à frente, malha fechada, c/dif. algoritmos ..	56
FIGURA 19 Simulação do impacto de cenários de dist.social - novos casos .....	58
FIGURA 20 Infográfico com os números de novos casos com o 14º dia anterior .....	60

## LISTA DE TABELAS

TABELA 1 Exemplo da série de dados obtida do SEADE-SP .....	29
TABELA 2 Exemplo dos dados de índice de distanciamento social do SIMI-SP .....	31
TABELA 3 Cenários hipotéticos de distanciamento social .....	38
TABELA 4 Parâmetros utilizados nos cenários novos casos de 7 dias.....	40
TABELA 5 Parâmetros utilizados nos cenários novos casos de 14 dias .....	40
TABELA 6 Parâmetros utilizados nos cenários óbitos de 7 dias .....	40
TABELA 7 Parâmetros utilizados nos cenários óbitos de 14 dias .....	40
TABELA 8 Melhores resultados do planejamento fatorial completo .....	42
TABELA 9 Algoritmos de treinamento utilizados nesta etapa do trabalho .....	55
TABELA 10 Painel teórico dos achados desse trabalho .....	64

## Sumário

RESUMO.....	5
LISTA DE ABREVIATURAS E SIGLAS .....	7
1. INTRODUÇÃO .....	11
1.1 PROBLEMA DE PESQUISA .....	12
1.2 OBJETIVOS .....	13
1.2.1 Objetivo Geral.....	13
1.2.2 Objetivos específicos.....	13
1.3 JUSTIFICATIVA .....	14
1.4 DELIMITAÇÃO DO TEMA.....	15
2 REFERENCIAL TEÓRICO .....	17
2.1 BREVE HISTÓRICO .....	17
2.2 MODELAGEM DA PANDEMIA.....	18
2.3 DISTANCIAMENTO SOCIAL E ISOLAMENTO SOCIAL .....	20
2.4 MODELO AUTORREGRESSIVO NÃO LINEAR C/ VARIÁVEL EXÓGENA ...	24
2.4.1 Formulação matemática do modelo NARX .....	25
3 MATERIAIS E MÉTODOS.....	28
3.1 OBTENÇÃO DOS DADOS .....	29
3.2 TRATAMENTO DOS DADOS .....	31
3.3 ARQUITETURA DA REDE NEURAL NARX.....	34
3.4 CALIBRAÇÃO DO MODELO NARX.....	35
3.5 CENÁRIOS DE DISTANCIAMENTO SOCIAL.....	37
3.7 FUNÇÃO DE ATIVAÇÃO .....	41
4 RESULTADOS E DISCUSSÃO .....	42
4.1 RESULTADOS DO IMPACTO DO DISTANCIAMENTO SOCIAL .....	57
5 CONCLUSÕES .....	62
5.1 SUGESTÕES DE TRABALHOS FUTUROS .....	65
Referências bibliográficas .....	66

## 1. INTRODUÇÃO

O controle sobre doenças virais infecciosas e contagiosas sempre foi considerado um grande problema para a saúde pública mundial (GAO, 2018), mas a pandemia demandou uma série de medidas, que foram recomendadas pela Organização Mundial de saúde (WHO - *World Health Organization*), e seguidas por boa parte dos governantes mundiais.

O distanciamento e isolamento social aparecem como uma das principais medidas (AQUINO et al., 2020). O distanciamento social consiste na diminuição da interação entre as pessoas, quando existe uma transmissão comunitária do vírus, podendo ser efetivada pelo fechamento de escolas, espaços públicos e cancelamento de eventos, com o objetivo de reduzir a velocidade da transmissão. O isolamento social, por sua vez, é uma medida que separa as pessoas infectadas das pessoas não infectadas e pode ocorrer em domicílio ou em hospitais (FREEDMAN, 2020).

O distanciamento e o isolamento social passaram a fazer parte da rotina de pessoas de todo o mundo, assim como a higienização constante das mãos e uso de máscaras para conter o aumento do número de novos casos e óbitos. Houve a suspensão do funcionamento de escolas, comércios e locais propensos à aglomerações, permitindo que somente funcionasse o que fosse considerado essencial. No Brasil, com o objetivo de garantir estas medidas foi necessário a imposição da lei 13.979/2020, conforme salientam Garcel e Souza Netto (2020), que poderia restringir o trânsito de cidade, bloquear estradas e garantir que somente o comércio essencial, como farmácias e supermercados funcionasse, o que é conhecido como *lockdown*. Já a lei 14.019/2020 tornou obrigatório o uso de máscaras de proteção individual.

Estas medidas de contenção, em geral, e o distanciamento social, em particular, objetivaram a mitigação da pandemia de COVID-19 por meio da redução no número de novas infecções. Entretanto, os seus reais efeitos quantitativos no número de novos casos e óbitos pela doença ainda estão sendo conhecidos, bem como as suas consequências sociais e econômicas

A relação do distanciamento social com a dinâmica da pandemia foi pesquisada em vários trabalhos, envolvendo desde a pesquisa de modelos epidemiológicos compartimentais até o tratamento de imagens com o reconhecimento de padrões para

identificação de aglomerações em espaços públicos, os quais são discutidos no referencial teórico. No entanto, uma análise quantitativa da relação entre o índice de distanciamento e o número de novos casos e óbitos é ainda incipiente. Este trabalho realiza uma investigação quantitativa do impacto do índice de distanciamento social no número de casos e óbitos pela COVID-19 ao propor um modelo autorregressivo fundamentado em redes neurais, o qual se caracteriza por usar valores passados e futuros das séries temporais correspondentes. O modelo proposto considera como uma série externa adicional o índice de distanciamento social, possibilitando relacionar esse índice às variáveis da dinâmica da doença (novos casos e óbitos).

Esta dissertação está organizada da seguinte forma: ainda neste capítulo apresenta-se o problema de pesquisa, os objetivos gerais e específicos, a justificativa e a delimitação do tema. No capítulo 2, é apresentado um breve histórico sobre quarentenas e um referencial teórico sobre aplicações com inteligência artificial no cenário da pandemia. No capítulo 3 é apresentada a metodologia e os instrumentos de pesquisa. O capítulo 4 é dedicado aos resultados dos experimentos. No capítulo 5 apresenta-se a conclusão do trabalho e as sugestões para trabalhos futuros.

## 1.1 PROBLEMA DE PESQUISA

A diminuição da taxa de transmissão é um dos primeiros objetivos das medidas de contenção, como o distanciamento social. Aquino et al. (2020), através de uma revisão narrativa sistematizou evidências de atos do poder público, onde sugere-se que o distanciamento social é efetivo. No entanto, algumas questões sobre o impacto do distanciamento social na sociedade ainda não são completamente respondidas. O distanciamento social parte do princípio de que toda população possui um local de moradia fixa, e que ainda pode usufruir das mínimas condições de subsistência caso seja necessário manter essa medida por um longo período (PIRES, 2020). Neste contexto, a contaminação estaria também relacionada a outras variáveis, como a condição social dos indivíduos.

Por outro lado, não se conhece o impacto quantitativo que variações no índice de distanciamento podem gerar no número de casos e óbitos pela doença, o que pode ser muito importante para orientar medidas restritivas em novos cenários de pandemia. Especialmente considerando as recentes previsões sobre o aumento na frequência das pandemias que a humanidade deverá enfrentar, é fundamental

aprender lições que possam ser usadas em futuras ocorrências desses eventos (THACKER et al., 2021; POKHREL, 2022; COCCIA, 2022).

Com objetivo de procurar investigar estas questões, um conjunto de experimentos computacionais foram realizados com modelos não lineares de forma a responder à pergunta que norteia a execução deste trabalho:

### **Qual o impacto do distanciamento social no número de novos casos e óbitos durante a pandemia de COVID-19?**

#### **1.2 OBJETIVOS**

Com o propósito de responder à questão de pesquisa foram elaborados os objetivos gerais e específicos, conforme descritos a seguir.

##### **1.2.1 Objetivo Geral**

O objetivo geral deste trabalho é avaliar o impacto do distanciamento social no número de novos casos e número de óbitos de COVID-19 no Estado de São Paulo, Brasil.

##### **1.2.2 Objetivos específicos**

Os objetivos específicos do trabalho são:

- Criar um modelo autorregressivo não linear que relacione a variável distanciamento social ao número de novos casos e número de óbitos de COVID-19.
- Avaliar o comportamento do número de novos casos e óbitos por meio da simulação de diferentes cenários hipotéticos com variações nos níveis de distanciamento social.
- Avaliar e quantificar as incertezas nas previsões do modelo por meio do cálculo do erro amostral dos resultados obtidos de um número pré-determinado de replicações.
- Avaliar resultados de previsão um passo de tempo à frente e múltiplos passos à frente para horizontes de previsão de 7 e 14 dias.

- Analisar a influência dos hiperparâmetros do modelo - número de camadas e neurônios por camada, algoritmo de treinamento e o atraso na série temporal, na acurácia da previsão.

### 1.3 JUSTIFICATIVA

A pandemia de COVID-19 atingiu a população mundial criando uma crise sanitária de grandes proporções. Segundo Werneck e Carvalho (2020), a sociedade precisou resolver um problema onde o conhecimento científico era insuficiente e a alta velocidade de disseminação do vírus gerava incertezas de quais seriam as melhores estratégias para enfrentamento da pandemia.

Além das mortes, as implicações econômicas dessa pandemia foram significativas. Segundo Houvèssou et al. (2020) a maioria dos países afetados bloquearam suas fronteiras e ordenaram o fechamento de fábricas, restaurantes, *shoppings* e diversas outras atividades econômicas.

Dessa forma, pesquisas continuam sendo realizadas para entendimento dos fatores que podem interferir na dinâmica da disseminação do SARS-Cov-19, bem como para ampliar a compreensão dos diversos fatores que podem ser úteis em futuras situações semelhantes. Dentre elas, as simulações computacionais são uma possibilidade para entendimento dos desdobramentos da pandemia, como aquelas que pesquisam o impacto do distanciamento social.

Nesse contexto, pesquisar o impacto do distanciamento social na pandemia e quantificar limiares de um nível de distanciamento social que possam, ao mesmo tempo, salvar vidas e impedir um colapso econômico, são ações importantes para definição de medidas restritivas de circulação de pessoas.

Sendo o distanciamento social parte importante da estratégia no combate à proliferação do vírus, torna-se necessário simular qual seria o índice capaz de diminuir os novos casos de contaminação. Um índice baixo de distanciamento social poderia não conseguir a diminuição de novos casos, já um índice alto poderia comprometer as necessidades essenciais e gerar outros problemas sociais. Para isso, o estudo de técnicas, modelos matemáticos e ferramentas que possam ajudar no entendimento dos desdobramentos da pandemia tornam-se importantes, e surge a necessidade de uma abordagem que vá além de modelos estatísticos. Nesse contexto, o uso das



redes neurais para extrair os padrões dos dados da doença pode ser adequado, pois os dados são de natureza dinâmica (LIESTOL, 2022; KRATSHMER, 2006).

Parte-se, também, do pressuposto de que as relações entre variáveis biológicas são mais bem descritas por funções não lineares. Processos como crescimento, decaimento e mortalidade raramente são relacionados linearmente às variáveis explicativas (SCHABENBERGER; PIERCE, 2002, capítulo 5). De acordo com Mollenhoff e Dette (2019), os modelos de regressão são muito utilizados para descrever a relação entre uma e múltiplas variáveis e uma única resposta. Neste sentido, pode-se citar as redes neurais artificiais (RNAs), que são utilizadas para modelar relações envolvendo séries temporais complexas. Segundo Galvão et al. (1999), em função de sua estrutura não-linear, as redes possibilitam entender características mais complexas dos dados.

#### 1.4 DELIMITAÇÃO DO TEMA

O trabalho tem como proposta pesquisar a modelagem dos dados da pandemia de COVID-19 em um modelo autorregressivo não linear. Está fora do escopo desta pesquisa a comparação com outros modelos autorregressivos, bem como a modelagem e simulação da dinâmica de propagação da doença por meio de modelos compartimentais, como SIR, que avaliam a evolução temporal do número de indivíduos em diferentes estados da doença (KERMACK, MCKENDRIK, 1927). Assim, os modelos da epidemiologia para investigar, analisar e prever a dinâmica da pandemia são apenas citados para composição da plataforma teórica sobre o tema.

A simulação de diferentes níveis de distanciamento e seus respectivos efeitos pode auxiliar a tomar decisões em situações de pandemias futuras, as quais consigam conciliar questões sociais, médicas e econômicas, bem como avaliar quais teriam sido as consequências de diferentes níveis de distanciamentos social. No entanto, o modelo desenvolvido nesse trabalho não visa primordialmente estimar o número de novos casos e o número de óbitos em cenários futuros, mas sim relacionar a variável distanciamento social com as variáveis número de casos e óbitos. Nesse sentido, capturar o padrão das variações dessas variáveis é considerado mais importante que a precisão das estimativas realizadas.

Apesar da possível e provável relação entre o distanciamento social e as questões econômicas, políticas, culturais e históricas, o trabalho avalia apenas o impacto do distanciamento no número de casos e óbitos pela COVID-19. Acredita-se que a partir dessa análise seja possível avaliar a possibilidade de considerar outros patamares de níveis de distanciamento em relação aos observados durante a pandemia, bem como mensurar o impacto da redução desses níveis na evolução da doença.

Por fim, este trabalho de pesquisa não objetiva questionar e/ou criticar o distanciamento social como medida de contenção da pandemia de COVID-19.

## 2 REFERENCIAL TEÓRICO

### 2.1 BREVE HISTÓRICO

A definição de políticas para o distanciamento de pessoas para mitigação do contágio de enfermidades é um conceito antigo. O isolamento de pessoas com probabilidade de estarem contaminadas com alguma doença contagiosa é utilizado desde o século XIV, como cita Neto (2020). Na história, esta separação de pessoas é definida como quarentena, e era também aplicada a animais e bens em geral (PANZAC, 1986).

Em 1347, a Peste Negra assolou a Europa, dizimando entre 20% e 25% da população do continente, numa época em que ainda se conhecia muito pouco sobre doenças transmitidas por micro-organismos. Conti et al. (2004) cita que o medo e o desconhecimento levaram autoridades da época a tomar medidas extremas. A primeira quarentena foi oficialmente decretada pelo Reitor do Porto de Ragusa (atual Dubrovnik) que decretou 30 dias de isolamento para passageiros de navios e 40 dias para viajantes por terra. Visconde Bernardo de Reggio determinou que todo doente da “peste” fosse levado para lugares distantes da cidade e lá permanecesse até se recuperar ou morrer. Durante os próximos 100 anos diversas leis similares foram introduzidas em portos da Itália e da França.

A definição do conceito de quarentena, sob um ponto de vista epidemiológico-científico, ocorreu somente no século XIX. A aplicação da medida não era sistemática e capilar e não contribuía muito para epidemias que ocorriam, como a de cólera. A França já havia definido em 1834 um encontro para uma padronização da quarentena, mas foi apenas em 1851 que a primeira Conferência Internacional Sanitária ocorreu em Paris.

Medidas parecidas às que foram aplicadas na peste Negra foram usadas de forma similar para controlar a transmissão de tuberculose durante a Grande Praga Branca entre 1880 e 1930. Mesmo durante os anos 50, quando já existiam medicações para combate à doença, a prática da quarentena ainda era indicada.

Mais recentemente, em 2003, o mundo passa por epidemias causadas por coronavírus. Um modelo de monitoramento é criado na Europa para controlar epidemias por conta do grande desafio advindo da epidemia de SARS (síndrome respiratória aguda grave). A humanidade presenciaria também, em 2012, a epidemia

de MERS (síndrome respiratória do Oriente Médio), também causada por outro coronavírus. Assim, na falta de vacinas, medidas de prevenção foram adotadas, e novamente a quarentena foi uma solução encontrada para estas epidemias.

Observa-se assim, que o distanciamento de pessoas durante epidemias e pandemias foi amplamente utilizado ao longo dos séculos, e ainda é um tipo de medida indicada por diversas organizações internacionais.

De acordo com Freedman (2020), ainda que os termos quarentena, isolamento e distanciamento social se assemelhem, os mesmos se distinguem. Quarentena é a restrição do movimento de pessoas que podem ter sido expostas a uma doença contagiosa, já o distanciamento social envolve medidas restritivas que objetivam a diminuição da interação entre as pessoas, que podem estar infectadas, mas ainda não identificadas. O isolamento é a separação de pessoas doentes daquelas não infectadas.

Desta forma, este trabalho utiliza o termo distanciamento social, já que os dados aferidos e coletados para os experimentos se referem à movimentação de pessoas durante a pandemia, mas não as identifica como contaminadas ou não, e tampouco é conhecido se elas foram expostas ao vírus.

## 2.2 MODELAGEM DA PANDEMIA

A avaliação da dinâmica da pandemia é importante para dar suporte a ações para prever o aumento da utilização do sistema de saúde e, também, efeitos colaterais na economia, conforme é apontado por Li et al. (2020). O modelo SIR (Suscetíveis-Infectados-Recuperados), desenvolvido por Kermack e McKendrick (1927) foi usado para auxiliar no entendimento de epidemias, como as de rubéola, sarampo e varíola. Adaptações do modelo SIR foram aplicadas na pandemia de COVID-19, ajudando a compreender o comportamento do vírus (MAHAYANA et al., 2021 ; PERAKIS et al., 2022). Esses modelos são compartimentais, ou seja, a população é dividida em compartimentos que refletem o estado que os indivíduos se encontram no desenvolvimento da doença, como suscetíveis, infectados e removidos. Observa-se na literatura trabalhos com variações do modelo SIR, como em Quintero et al. (2021), que realizaram uma pesquisa com o modelo SEIRD (Suscetíveis, expostos, infectados, recuperados e mortos), onde foi proposto um modelo preditivo para as

variáveis do modelo baseado em seis variáveis adicionais: população, pessoas acima de 65 anos, índice de pobreza, taxa de morbidade, idade média e número de pessoas por quilômetro quadrado.

Além dos modelos compartimentais, verificou-se a importância de analisar o impacto dos testes de COVID-19 na dinâmica da pandemia, já que uma pessoa infectada poderia ser orientada ao isolamento social, neste sentido, Griette e Magal (2021), utilizam uma equação diferencial, a qual usa o número diário de testes como *input*. Também exploram a relação entre o número de testes realizados e a quantidade de novos casos. Ainda são simulados diferentes cenários, onde projeta-se o aumento do número de testes em 2,5, 10 e 100 vezes. Concluiu-se que, aumentando em 10 vezes o número de testes, haveria uma diminuição de indivíduos infectados, porém, após este limiar, não se notou grande diferença. Sendo assim, torna-se necessário encontrar uma estratégia otimizada, já que existem outras variáveis relacionadas, como o custo dos testes. A estratégia de testagem é uma importante ferramenta para o desenho de políticas de distanciamento social, outra estratégia utilizada na pandemia de COVID-19. Schimit (2020) analisa o impacto das características de distanciamento social na dinâmica populacional; estima o número de óbitos por COVID-19 por falta de infraestrutura de saúde e estuda combinações de recursos para evitar o colapso do sistema de saúde. No trabalho, um modelo SEIR (suscetível-exposto-infectado-removido) é descrito em termos de autômatos celulares probabilísticos e equações diferenciais ordinárias para a transmissão de COVID-19. O modelo é flexível para simular diferentes cenários de distanciamento social de acordo com as seguintes características: o dia de início para o distanciamento social após a primeira morte, o período de campanha de distanciamento social, e a porcentagem da população comprometida com a campanha. Os resultados mostraram que os esforços na campanha devem se concentrar tanto na taxa de distanciamento social como na sua duração.

Além destes modelos, segundo Ribeiro et al. (2020), uma das contribuições significativas para o enfrentamento da pandemia são aquelas que incorporam inteligência artificial, como as de aprendizado profundo para produzir modelos preditivos precisos.

Assim, vários outros estudos foram realizados. Chimmula e Lei (2020) propõem um modelo de previsão de COVID-19, no Canadá, utilizando LSTM (*long-short term memory*) o qual se baseia num modelo *Deep Learning*. A rede foi treinada com dados

até 31 de março de 2020, referentes a número de casos confirmados, mortes e pacientes recuperados. Para mitigar o ruído das séries foi aplicada uma transformada wavelet. O teste *Dickey Fuller* foi aplicado nos dados de entrada para verificar a estacionariedade da série temporal. Os dados foram então inseridos numa rede LSTM. Após simulações foi possível prever o índice de contaminação e o fim da pandemia com acurácia de 93,4%. Além destes modelos, a área de inteligência artificial na saúde contribuiu também no sentido de predizer diagnóstico de COVID-19 em pacientes sintomáticos, Moraes Batista et al. (2020), através de algoritmos de aprendizado de máquina treinados com dados de pacientes sintomáticos, utiliza variáveis preditoras como idade, sexo e resultados de hemogramas, já que exames do tipo RT-PCR (*real time protein C reative*) podem demorar dias, e um diagnóstico mais rápido pode ajudar na tomada de decisão, como o isolamento do paciente. O modelo foi treinado com 235 pacientes (102 positivos para COVID-19) e obteve AUC (*area under the ROC curve*) de 0,85. O objetivo do modelo não era dar um diagnóstico categórico para COVID-19, mas sim, através de uma probabilidade, possibilitar a tomada de medidas proativas.

Nota-se que as pesquisas utilizaram de diversas abordagens para prevenção, entendimento e predição de índices de contágio e mortes, para que as instituições públicas pudessem tomar medidas que mitigassem o efeito da pandemia. Segundo Wang et al (2020), boa parte dos estudos tiveram como objetivo a compreensão dos sintomas e a estimação dos períodos de incubação. Diversas pesquisas mostram que o período de incubação da infecção viral varia de 4 a 14 dias e muitos pacientes são assintomáticos.

Os trabalhos acima são apresentados para definição da plataforma teórica sobre o tema. O referencial teórico com modelos não lineares é discutido na seção 2.4.

## 2.3 DISTANCIAMENTO SOCIAL E ISOLAMENTO SOCIAL

A partir da observação do que ocorreu na China, que imediatamente aplicou restrições de viagens, circulação de pessoas e o monitoramento de sintomas entre viajantes, conforme cita Chinazzi et al. (2020), ficou evidenciado ao mundo a gravidade iminente do COVID-19. O distanciamento social foi uma importante

ferramenta e tornou-se uma das poucas medidas possíveis para autoridades em todo o mundo, que precisavam resolver uma crise sanitária global.

No Brasil, um país com imensas desigualdades sociais, onde as relações entre pobreza e saúde são percebidas no cotidiano da sociedade brasileira (PITOMBEIRA e OLIVEIRA, 2020) são necessárias medidas de proteção aos mais pobres, de forma a garantir a adesão às medidas de distanciamento social.

A introdução de diversas medidas de contenção de uma só vez com diferentes graus de adesão tornou um desafio avaliar a efetividade das medidas isoladamente. Na bibliografia encontramos diversos trabalhos que estudaram os impactos do distanciamento social logo no início da pandemia.

Um estudo em Wuhan, que utilizou dados de COVID-19 associados a informações obtidas de telefones celulares, apontou que a mobilidade das pessoas foi o principal fator de propagação do SARS-CoV-2, antes da implementação das medidas de isolamento. Nesse sentido, a redução da mobilidade de pessoas pode contribuir para retardar o pico da epidemia e evitar a transmissão para outros locais (KRAEMER et al., 2020 ; CHINAZZI et al., 2019 ; BOLDOG et al., 2020 ; LAU et al., 2020)

Ainda em Wuhan, Kucharski et al. (2020) avaliou a restrição de viagens entre dezembro de 2019 e fevereiro de 2020, onde foi verificada uma redução da transmissão, no final de janeiro de 2020. Foi estimado que o fechamento dos aeroportos na China, que ocorreu cerca de dois meses após o início da epidemia, promoveu um adiamento na ocorrência de novos casos fora de Wuhan, tanto no resto da China, quanto internacionalmente. Todavia, estimou-se que reduções de até 90% no número de voos só diminuiria o número de casos em outros países se a detecção precoce, o isolamento e mudanças comportamentais na população (como lavar as mãos, evitar aglomerações, entre outras) fossem implementados e incentivadas de forma conjunta (CHINAZZI et al., 2020).

Aquino et al (2020) classifica as medidas do distanciamento social em categorias, onde observa-se diversos desdobramentos como teletrabalho, proibição de eventos, suspensões de transporte internacional e interestadual, fechamento de unidades de ensino, proibição de frequentar parques e praias, dentre outras medidas impostas através de decretos. Nota-se que a sociedade civil se movimentou para um bem comum e seus impactos continuam até hoje, como exemplo, as empresas que ainda continuam em regime de teletrabalho. O trabalho conclui que as medidas,

principalmente aquelas que reduzem em pelo menos 60% os contatos sociais, têm o potencial de diminuir a transmissão da doença.

A efetividade das restrições de distanciamento social também foi pesquisada por Hou et al. (2020), onde são comparados diferentes cenários e através do SEIR e realizam uma previsão do número de infectados. O método analisou 11 estimativas através de modelos de regressão, comparando os resultados através do erro quadrático, onde o impacto do distanciamento social foi considerado importante na redução de novos casos e óbitos.

Em todo o mundo, políticas mais rígidas como o *lockdown* foram tomadas, segundo Ferguson et al (2020) o termo designa o banimento de qualquer atividade não essencial, inclusive o fechamento de instituições educacionais, de forma que as pessoas fiquem em casa, diminuindo ao máximo possível o contato social.

Observa-se, então, a necessidade do uso de tecnologia para mensurar a adesão ao distanciamento social. Neste sentido foram realizadas pesquisas de algoritmos de detecção de movimento de pessoas. Rezaei e Azarmi (2020) desenvolveram um modelo híbrido baseado em redes neurais de aprendizado profundo utilizando câmeras comuns de segurança. O modelo proposto é uma combinação entre uma rede neural de aprendizado profundo e algoritmos de *tracking* de movimento. O método foi aplicado em um *dataset* com 7530 *frames*, detectando e estimando distância em aproximadamente 150.000 pessoas. Esta pesquisa surge como uma opção de monitoramento do distanciamento social em espaços públicos e terminais de transporte urbano, onde são utilizados equipamentos que já existem nestas estruturas. Acrescidos de tratamento de imagens, esta análise pode ajudar na avaliação de medidas preventivas.

Ainda no sentido de entender o impacto do distanciamento social na dinâmica da pandemia de COVID-19, encontram-se pesquisas baseadas em redes neurais artificiais que realizam tratamento de imagens, nos quais, através de câmeras, consegue-se mapear se há ou não aglomerações. Pode-se observar pesquisas deste tipo em Tanwar et. al. (2021) e Niu et al.(2021).

Também se observa propostas com uso de IoT (Internet das coisas) para detecção de aglomerações baseada em Inteligência Artificial, como em Fedele e Merenda (2020).



Pesquisas também foram realizadas de forma a entender o distanciamento social indiretamente, como exemplo, através de perfil de consumo de energia elétrica, como em Airam et al. (2021).

Ainda se observa o desenvolvimento de sistemas de avisos para violações do distanciamento social, como pesquisou Dongfang et al. (2021). Muitos estudos aproveitaram trabalhos anteriores e adaptaram para o contexto da pandemia, de forma que pesquisas sobre otimização de caminhos de saída de emergência em locais de grande volume de pessoas, puderam contribuir no cenário da pandemia de COVID-19.

A utilização de tecnologias emergentes como aquelas baseadas em redes *wireless* e Inteligência Artificial podem permitir e até impor o distanciamento social. Nguyen et al. (2020) aponta que a informação sobre a quantidade de pessoas em um determinado lugar, que pode ser informada através de aplicativos de celular, pode orientar as pessoas a evitarem espaços assim, e por consequência, diminuir as chances de uma possível contaminação.

Araujo et al. (2021) propõe um modelo baseado numa arquitetura de rede neural, mensurando a acurácia através do MAPE (erro médio absoluto percentual). A pesquisa apresentou previsões altamente correlacionadas aos valores reais da série temporal, com desvios percentuais expressivamente baixos.

Lounis et al. (2021) faz uma pesquisa com modelos preditivos estimando números de novas contaminações e mortes, comparando modelos com equações diferenciais e modelos baseados em redes neurais artificiais. Observa-se que este último modelo consegue um coeficiente de determinação  $R^2 = 0.9999$ , os dados foram validados em um período de 26 dias. Todos os dados foram baseados em PCR positivos reportados, onde o número de contaminados pode ser muito maior.

Portanto, trabalhos recentes abordaram a influência de medidas de distanciamento social na dinâmica da pandemia. São escassos, no entanto, resultados quantitativos da medida do impacto dessas medidas e, em especial, na possibilidade de definição de um limiar de segurança para o índice de distanciamento. Adicionalmente, destaca-se que muitos dos modelos da literatura, em especial àqueles baseados no modelo SIR e suas variantes, usam parâmetros que demandam informações sobre a propagação da doença que não são completamente conhecidos no caso da COVID-19 como, por exemplo, os parâmetros de movimentação e número de contatos dos indivíduos da população. Tais limitações reforçam a necessidade do

desenvolvimento de modelos que utilizam diretamente, e apenas, os dados disponíveis sobre as variáveis de interesse como é o caso do modelo autorregressivo não linear apresentado na próxima seção.

## 2.4 MODELO AUTORREGRESSIVO NÃO LINEAR COM VARIÁVEL EXÓGENA

O Modelo Autorregressivo não linear com variável exógena (NARX) é um tipo de rede neural que pode modelar problemas não lineares entre variáveis de entrada e saída em séries temporais. Este tipo de modelo de regressão é muito utilizado para predição de valores, descrevendo a relação entre uma variável resposta  $Y$  e uma variável explicativa  $X$ . O modelo é linear se a quantidade de interesse, geralmente a média de  $Y$ , é função linear dos parâmetros, caso contrário é não linear (ZEVIANI et al., 2013). No caso autorregressivo, as variáveis explicativas e resposta são as mesmas e seus valores passados são usados para estimar valores futuros dessa variável.

A NARX possui duas partes: uma rede neural autorregressiva e uma série exógena. A série externa ou exógena são variáveis que podem influenciar as saídas. Este tipo de rede neural é frequentemente usado para sistemas dinâmicos não lineares. As aplicações deste tipo de rede são diversas, e pesquisas mostram diversas áreas que se utilizaram da mesma.

Aplicações com NARX foram utilizadas para previsão de níveis de reserva em uma represa, como pesquisou Supratid et al. (2017). Com dados históricos de 1984 a 1999, foi buscado uma projeção para o período de 2010 a 2099, a variável externa utilizada foi a precipitação perto da represa. Utilizando o RMSE (*root mean squares error*), a NARX obteve os melhores resultados, mostrando que neste cenário, uma modelagem utilizando uma série externa, aqui representada pelo índice de precipitação, melhorou a predição dos valores.

Na área de manutenção de sistemas mecânicos, Nistane (2019) pesquisou um prognóstico de degradação de rolamentos, onde a previsão de manutenção poderia evitar falhas em máquinas rotativas. Foram utilizados como dados históricos um sinal de vibração, e estes relacionados a falhas do rolamento. Em sistemas deste tipo, o atrito entre as peças degrada o equipamento, sendo assim, os dados são coletados através de um sistema de medição de torque. A proposta principal do trabalho é criar um indicador de saúde do sistema rotativo. Nessa proposta, a entrada foram os dados

de medição da vibração e após treinamento da rede, obteve-se como saída três conjuntos de dados, sendo: 1 – funcionamento normal, 2 – degradação e 3 – falha do sistema. Este modelo de *machine learning* relacionou as variáveis de entrada a um *status* de saúde do sistema, onde no futuro, a partir de medidas do torque, seja possível prever uma possível manutenção do sistema. As estimativas foram obtidas através do MSE (*mean squared error*). Para fins de comparação, o trabalho realizou o experimento através de uma rede neural recorrente e uma NARX, que conseguiu os dados mais precisos de degradação dos rolamentos e colaborou para definição do indicador de saúde.

Observa-se aplicação do modelo NARX também na economia, uma das áreas com mais volatilidade, onde previsões mais precisas podem orientar investimentos, gastos e aplicações financeiras. Neste sentido Leon e Ortega (2018) pesquisaram a predição do ISE - um indicador de atividade econômica de curto prazo na Colômbia. Utilizando dados coletados de pagamentos realizados entre indivíduos, empresas e governo central como dados de entrada, a NARX contribuiu para uma predição justa dois meses antes dos *releases* financeiros.

A economia possui indicadores com alta volatilidade e previsões mais acuradas orientam o poder público na tomada de decisões, investidores e público em geral para decisões econômicas.

A diversidade de áreas utilizando a NARX é ampla e sua aplicação pode ser importante para pesquisas epidemiológicas, como é proposto neste trabalho.

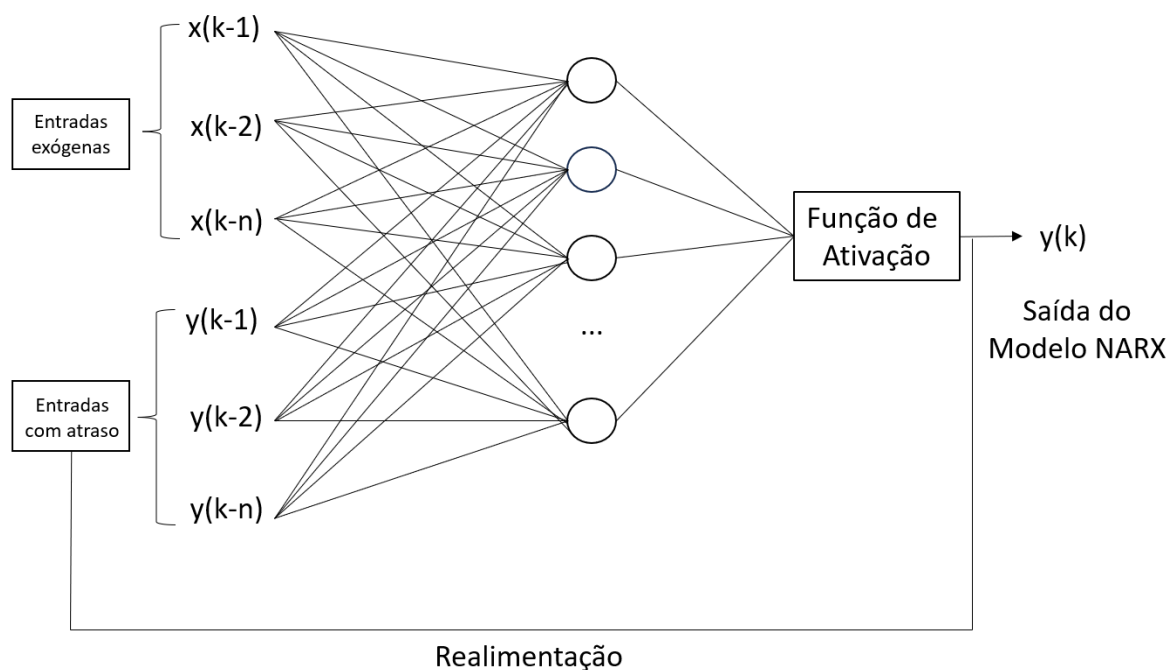
#### 2.4.1 Formulação matemática do modelo NARX

O modelo NARX possibilita a previsão de valores futuros de uma série temporal utilizando os dados históricos, por meio de um mecanismo de realimentação, no qual um valor previsto pode servir de entrada para novas previsões. Primeiramente, a rede é treinada com os valores da série temporal, para assim garantir maior precisão no treinamento, o que se define por *loop* aberto. Em um segundo momento, os valores previstos são utilizados como novas entradas para a rede – *loop* fechado.

Esses sistemas do tipo *loop* fechado, conforme pode ser observado na Figura 1 tentam emular recursivamente o comportamento dinâmico do sistema que gerou as séries temporais não lineares, nas quais o horizonte de precisão pode ser infinito. A

resposta prevista será gerada utilizando estimativas e acumulando o erro de previsão em cada laço (CHATTERJEE et al., 2012).

Figura 1: Rede neural artificial NARX.



Fonte: Autor

Uma série temporal é qualquer conjunto de observações organizadas no tempo, geralmente as medições são feitas em tempos espaçados uniformemente como, por exemplo, número de novos casos e mortes por uma doença, o índice de poluição, valores mensais de temperatura, valores diários de consumo de eletricidade, índices diários de bolsa, entre outros.

Na modelagem de séries temporais, um modelo NARX é um modelo autorregressivo não linear que possui entradas exógenas e tem sua arquitetura baseada na rede Perceptron Multi-camadas, que recebe entradas atrasadas no tempo. O modelo relaciona o valor atual de uma série temporal  $\{y(t)\}$  a valores passados da mesma série  $\{y(t-1), y(t-2), \dots, y(t-m)\}$ , e valores atuais e passados da série exógena  $\{x(t), x(t-1), x(t-2), \dots, x(t-n)\}$ , que passa então a influenciar a série de interesse. Os atrasos na variável principal  $y$  e na variável exógena  $x$ , denotados por  $m$  e  $n$ , respectivamente, podem ser supostos iguais e representados pela letra  $d$  como sendo o parâmetro de atraso de tempo do modelo. Além disso, o modelo contém um

termo de erro  $\varepsilon$  relacionado ao fato de que o conhecimento de outros termos não permitirá que o valor atual da série temporal seja previsto com exatidão.

Assim, o valor da série  $y$  no instante de tempo  $t$  é representado conforme definido na Equação (1) para uma variável exógena  $x$ ,

$$y(t) = F(y(t-1), y(t-2), \dots, y(t-d), x(t-1), x(t-2), \dots, x(t-d)) + \varepsilon \quad (1)$$

na qual  $F(y, x)$  é uma função não linear que relaciona as observações passadas de  $x$  e  $y$  com valores futuros da variável que se deseja prever,  $y$ .

A previsão de múltiplos passos à frente ( $y(t+1)$ ,  $y(t+2)$ , ...) pode ser realizada usando o valor predito  $y(t)$ , no passo de tempo  $t$ , como entrada para previsão do valor no próximo passo de tempo  $t+1$ , como representado pela equação (2). Importante observar que, nesse caso, é necessário conhecer o valor da variável exógena no passo de tempo  $t$ . Computacionalmente, essa operação é obtida fechando o *loop* da rede neural artificial (malha fechada) como ilustrado na Figura 1.

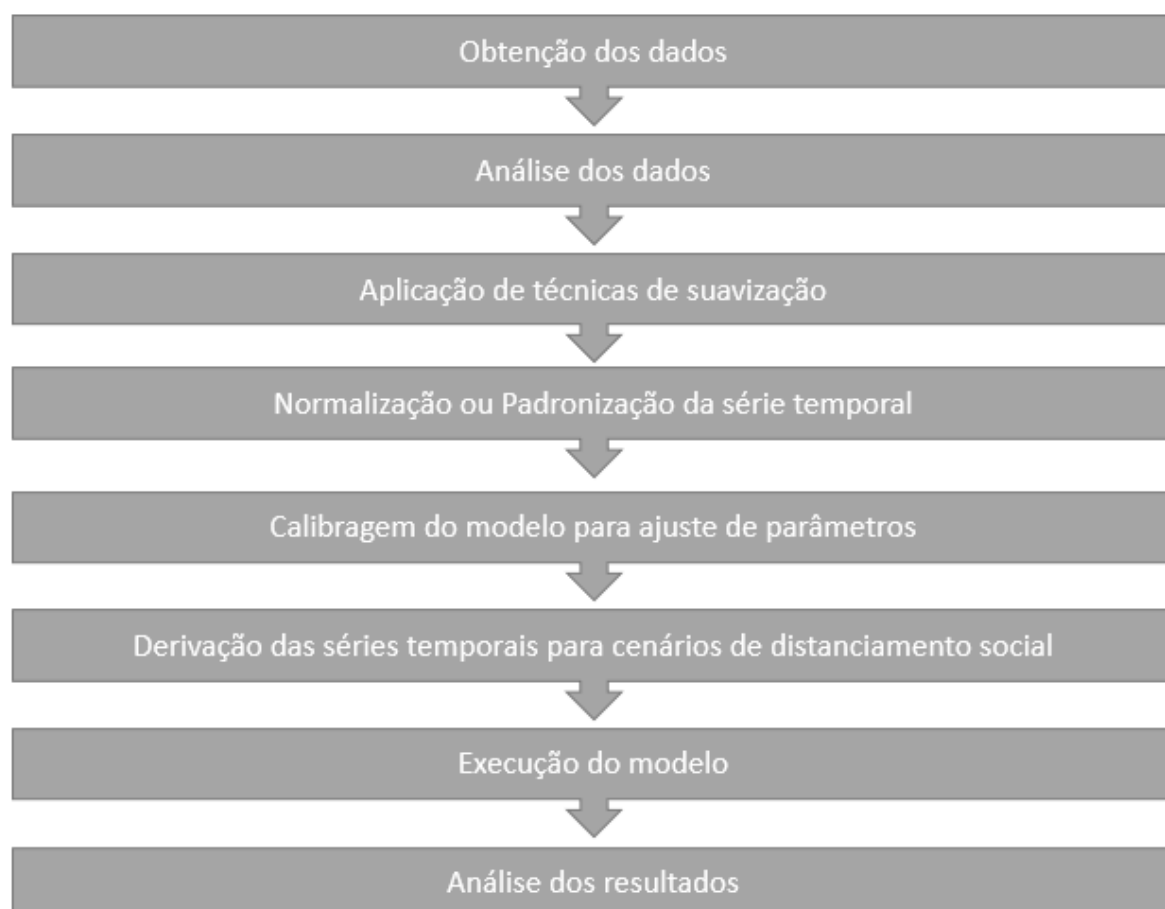
$$y(t+1) = F(y(t), y(t-1), \dots, y(t-d+1), x(t), x(t-1), \dots, x(t-d+1)) + \varepsilon. \quad (2)$$

### 3 MATERIAIS E MÉTODOS

Este capítulo descreve os materiais e métodos utilizados na realização dessa pesquisa, a qual pode ser caracterizada como uma pesquisa aplicada com abordagem experimental. A característica experimental é oriunda da manipulação direta das variáveis relacionadas ao objeto de estudo, a saber o número de novos casos e óbitos pela COVID-19 e o índice de distanciamento social. A busca pela quantificação da relação entre essas variáveis define o aspecto de pesquisa experimental (ANDRADE, 2017).

A realização dos experimentos seguiu as etapas de acordo com a Figura 2. O planejamento foi organizado com apoio de planilhas eletrônicas no *Microsoft Excel*, e os experimentos com o modelo NARX foram então processados no *Matlab* com posterior realização de análises comparativas.

Figura 2: Etapas dos procedimentos experimentais.



Fonte: Autor

### 3.1 OBTENÇÃO DOS DADOS

A criação de um modelo que associe o número de novos casos e óbitos por COVID-19 com o índice de distanciamento social requer a obtenção das respectivas séries de dados.

Os dados referentes ao número de casos e óbitos foram obtidos por meio da Fundação Sistema Estadual de Análise de Dados do estado de São Paulo (SEADE), os quais estão disponíveis em <https://www.seade.gov.br/coronavirus/>. A série de dados relaciona data, total de casos acumulados, casos por dia e óbitos, conforme ilustrado na Tabela 1, como um exemplo para o período de cinco dias entre 08/06/2020 e 12/06/2020. Todos os dados foram obtidos em formato CSV (*Comma separated values*).

Tabela 1: Exemplo da série de dados obtida do SEADE-SP.

Data	Total de casos	Casos por dia	Óbitos por dia
08/06/2020	144593	1520	43
09/06/2020	150138	5545	334
10/06/2020	156316	6178	340
11/06/2020	162520	6204	283
12/06/2020	167900	5380	223

Fonte: SEADE-SP

O próximo passo foi a obtenção dos dados de distanciamento social. Foram constatadas, por meio da revisão bibliográfica, poucas referências sobre como aferir a adesão ao distanciamento social. As opções pesquisadas foram aquelas relacionadas à tecnologia utilizada em aparelhos de celular utilizando GPS (*global positioning system*). O *Google* publicou dados de distanciamento social no *COVID-19 Community Mobility Report*, estes trazem informações sobre mobilidade em locais de varejo e lazer, mercados e farmácias, parques, estações de transporte público, espaços de trabalho e residências. A publicação dos dados teve grande preocupação com o anonimato, e não foram disponibilizados nos dias e locais nos quais não houve movimentação suficiente. A série de dados do *Google* não foi considerada para as próximas etapas da pesquisa, pois tem restrições e considera a movimentação apenas em lugares específicos, como ilustrado na Figura 3.

Figura 3: Exemplo de dados obtidos do Google COVID-19: Relatório de mobilidade da comunidade. Exemplo para cidade de Campinas, São Paulo.

### Campinas



Fonte: Google

Outras formas de medir o distanciamento social foram pesquisadas, como os dados do Sistema de Monitoramento Inteligente de São Paulo (SIMI-SP), viabilizado por meio e acordo com as operadoras de telefonia móvel Vivo, Claro, Oi e Tim, através da ABR (Associação Brasileira de Recursos em Telecomunicações) e do IPT (Instituto de Pesquisas Tecnológicas), onde são fornecidos dados de distanciamento social.

De acordo com as operadoras, o índice de distanciamento social é baseado na localização obtida pelas antenas de celulares (Estações Rádio Base – ERBs), as quais definem uma referência para o lugar onde o celular passou a noite (entre 22h00 e 2h00). Durante o dia, um celular que tenha se deslocado desta marcação (com erro aproximado de 200m), passa a considerar esta informação na composição do índice de distanciamento social. O índice foi atualizado diariamente, sempre mostrando os valores referentes ao dia anterior. Este espaço de tempo ocorre em função do trabalho das operadoras para agregar e retirar qualquer informação que possa identificar o usuário.

Os dados foram coletados do sítio público do SIMI-SP disponível em <https://www.saopaulo.sp.gov.br/coronavirus/isolamento/>, entre os dias 26 de fevereiro



de 2020 e 13 de julho de 2021 referentes ao Estado de São Paulo/SP. O índice de adesão ao distanciamento social relaciona a cidade, a data e o índice, representado por um percentual.

Tabela 2: Exemplo dos dados de índice de distanciamento social do SIMI-SP

Município	Código Município IBGE	População estimada (2020)	UF1	26/02/20	27/02/20	28/02/20	29/02/20	01/03/20	02/03/20	03/03/20
AMERICANA	3501608	242018	SP	36%	34%	34%	35%	37%	34%	33%
AMPARO	3501905	72677	SP	33%	33%	33%	35%	36%	36%	36%
ANDRADINA	3502101	57202	SP	37%	36%	36%	36%	37%	39%	38%

Fonte: SIMI-SP

A Tabela 2 mostra o índice de distanciamento social por data. Como exemplo, na cidade de Americana, a adesão ao distanciamento social em 26/02/20 foi de 36% (de acordo com a metodologia escolhida para obtenção dos dados através das antenas de celulares). Assim, estes dados foram eleitos para as próximas etapas da pesquisa.

Os dados obtidos têm relação com a movimentação de pessoas, considerando seu deslocamento através daquele obtido pelos dados captados de telefones celulares. A revisão bibliográfica apresentou poucas formas de mensurar a circulação de pessoas, e as medidas do SIMI são as que mais se aproximam do conceito de distanciamento social. Desta forma, o uso destes dados pressupõe que o deslocamento de uma pessoa durante o dia é uma medida que infere o distanciamento social em uma cidade.

Os dados referentes a número de casos, óbitos e índice de distanciamento social foram tratados e unificados em apenas um arquivo utilizando o *Microsoft Excel*, resultando em três séries temporais de 500 amostras, uma para cada variável.

### 3.2 TRATAMENTO DOS DADOS

Os dados coletados, conforme descrito na seção anterior, foram pré-processados por meio de uma combinação das operações de média móvel e normalização, em um primeiro experimento, e média móvel e padronização nos experimentos finais.

Os estados brasileiros, em seus boletins epidemiológicos, definiram conceitos e definições para o monitoramento de COVID-19. Um destes conceitos é que a média

móvel deve considerar uma janela de tempo de 7 dias, conforme pode ser observado, como exemplo, no Boletim Epidemiológico da Secretaria de Saúde do Distrito Federal.

Resultados da aplicação da média móvel aos dados podem ser observados nas Figuras 4, 5 e 6 para o número de novos casos, número de óbitos e índice de distanciamento social, respectivamente. Tais resultados foram obtidos por meio da função `movmean(y, 7)` no Matlab, sendo  $y$  a variável de interesse.

Figura 4: Dados originais da série temporal do número de novos casos de COVID-19 e Média Móvel dos últimos 7 dias.

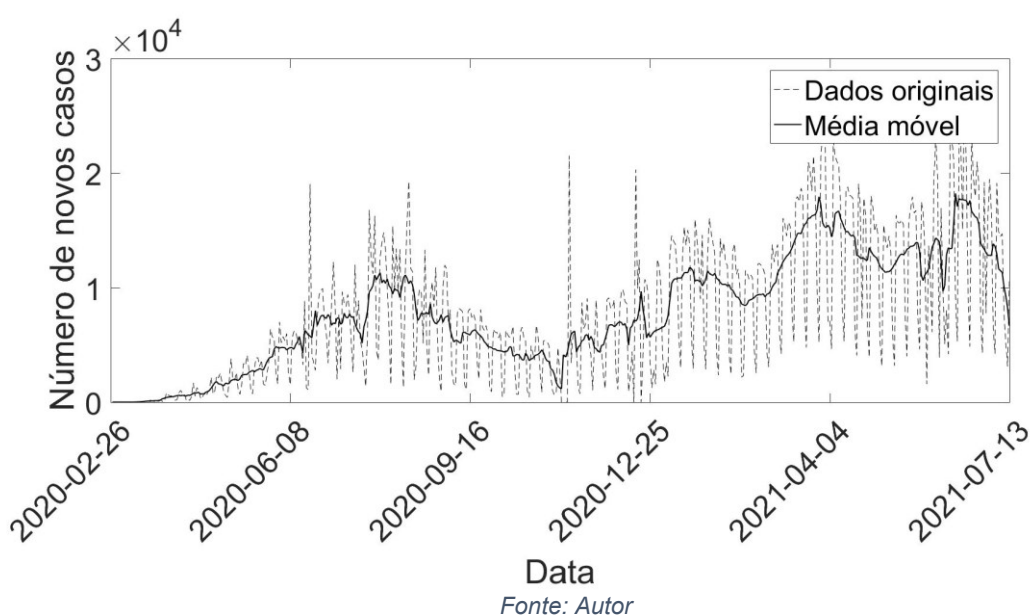


Figura 5: Dados originais da série temporal de número de óbitos por COVID-19 e média móvel dos últimos 7 dias.

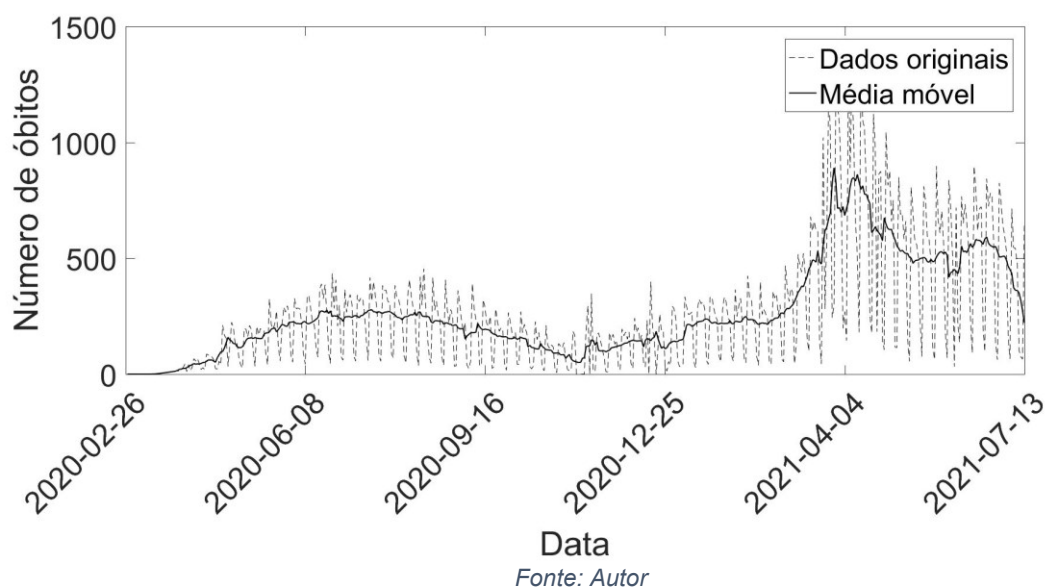
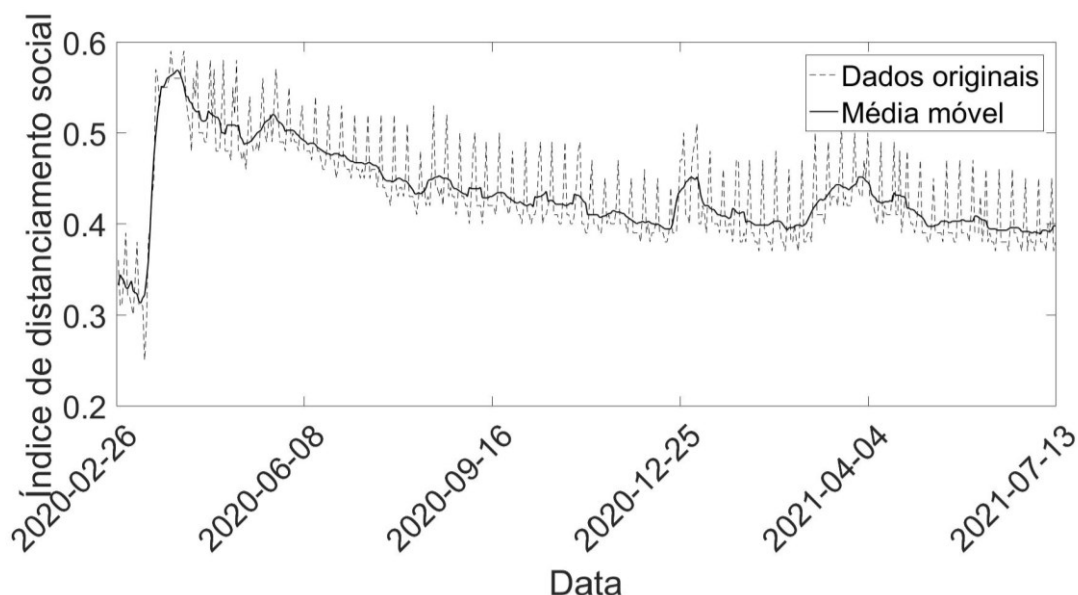


Figura 6: Dados originais da série temporal de índice de distanciamento social e média móvel dos últimos 7 dias.



A utilização de uma janela de 7 dias para média móvel também se deveu ao fato de ser a metodologia utilizada pelas autoridades para avaliar e divulgar os resultados da COVID-19, por ser uma alternativa simples para suavizar as oscilações e ajudar no entendimento do avanço da pandemia. A média móvel dos últimos 7 dias também foi adotada pelo Consórcio formado pelas empresas de comunicação UOL, G1, O Globo, O Estado de São Paulo e Folha de São Paulo.

Após a suavização dos dados com a média móvel, dois conjuntos de experimentos foram realizados. Primeiramente, os dados foram normalizados de forma que os valores ficassem distribuídos entre 0 e 1, como é usual em pesquisa com redes neurais. Nesse caso, os valores da série temporal são divididos pelo valor máximo da série no período de análise. A normalização de dados é um processo para colocar os valores em uma escala comum, sem que haja distorções ou perda de informação. De acordo com Huang et al.(2020) essa operação é recomendada, em geral, em situações nas quais os dados originais não satisfazem uma distribuição normal e apresentam assimetrias significativas. Adicionalmente, são muito sensíveis à presença de outliers.

Em um segundo conjunto de experimentos, realizados para efeitos de comparação e buscando melhorar os resultados obtidos na etapa anterior, os dados passaram por uma operação de padronização, a qual objetiva transformar a série de

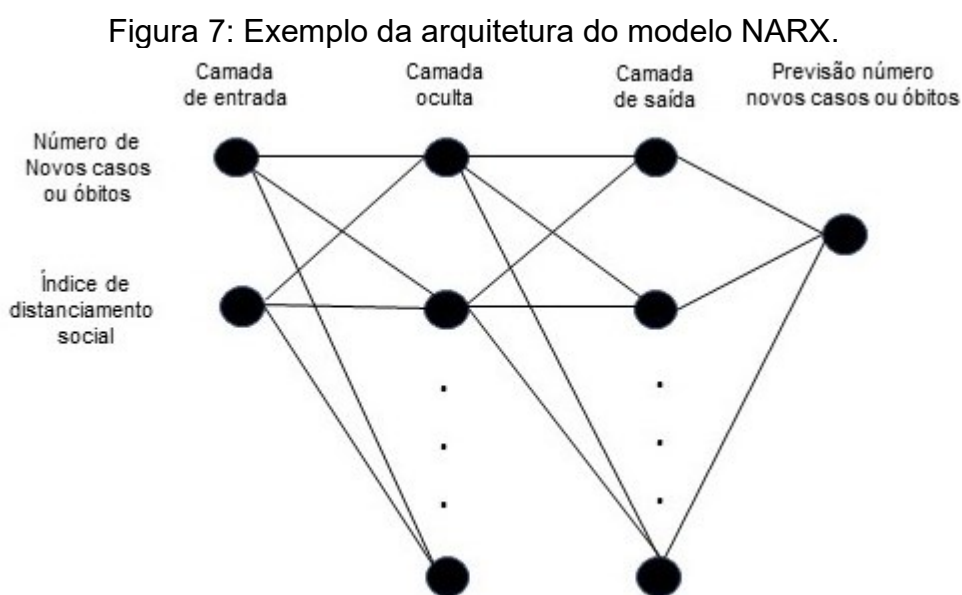
dados para que se obtenha média igual a zero e desvio-padrão igual a 1. A operação de padronização é obtida utilizando a fórmula do z-score definida na Equação (3) e, em geral, é considerada mais adequada quando os dados se aproximam de uma distribuição normal e há presença de *outliers* (HUANG et al., 2020).

$$z = \frac{y - \mu}{\sigma}, \quad (3)$$

na qual  $y$  é a variável a ser padronizada, com média  $\mu$  e desvio-padrão  $\sigma$ .

### 3.3 ARQUITETURA DA REDE NEURAL NARX

A camada de entrada da rede neural NARX recebe os dados de novos casos (ou os dados de óbitos) e o índice de distanciamento social, em atrasos de tempo pré-definidos no planejamento de testes. Os dados então são propagados em direção à(s) camada(s) oculta(s) para processamento. Cada camada oculta (*Hidden layer*) contém funções de ativação, calcula os pesos das variáveis e, assim, verifica os efeitos dos preditores (valores passados das séries) sobre as variáveis de interesse (valores futuros de novos casos e óbitos). Na camada de saída, o processo de predição é finalizado e os resultados são apresentados com estatísticas de erro. Um exemplo da arquitetura do modelo é apresentado na Figura 7.

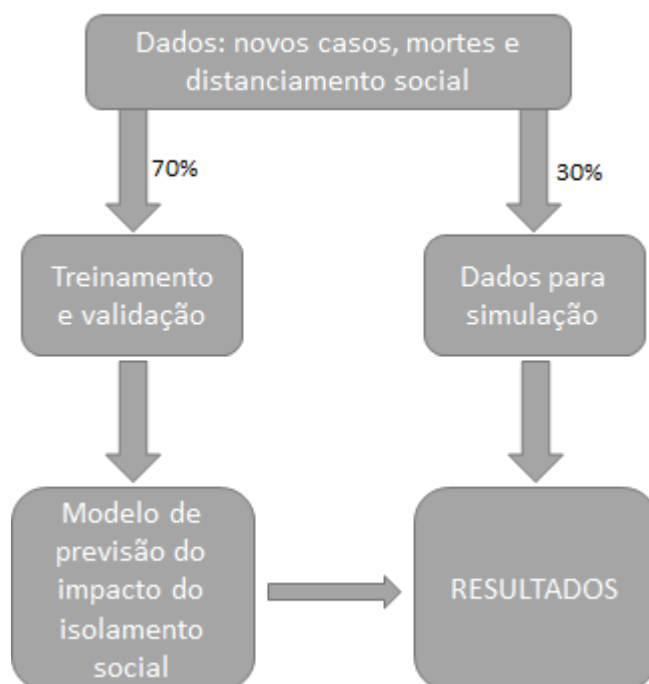


Na Ilustração simplificada da rede NARX apresentada na Figura 7, o número de entradas da rede é proporcional ao atraso  $d$  utilizado nas variáveis principal e exógena. Por exemplo, nas simulações com  $d=7$  tem-se 14 entradas no modelo, sendo sete valores de novos casos (ou óbitos) e sete valores de distanciamento social. Neste trabalho, os atrasos utilizados para as duas variáveis são sempre iguais em função das características do problema e da ferramenta computacional adotada.

### 3.4 CALIBRAÇÃO DO MODELO NARX

Nesta etapa do experimento, objetivou-se a calibração dos parâmetros da rede NARX. Os dados foram divididos em três conjuntos distintos para treino, validação e teste, respectivamente. Como é usual, foi definido 70% dos dados para o treino e 30% para validação e testes (15% e 15%), conforme pode ser observado na Figura 8. A arquitetura utilizada dividiu os dados randomicamente, de forma que os dados de treinamento sejam sempre diferentes nas 30 replicações realizadas em cada caso, nas mesmas condições.

Figura 8: Esquema de treinamento da rede neural



Fonte: Autor

Foi realizado um planejamento de experimentos para calibração de parâmetros no modelo NARX. O planejamento de experimentos é um conjunto de técnicas estatísticas aplicadas ao planejamento, condução, análise e interpretação de testes controlados para definir os fatores que influenciam os valores de um parâmetro ou grupo de parâmetros. Segundo Kalil et al. (2000), os planejamentos verificam o efeito das variáveis independentes e estimam o efeito das interações entre estas sobre a variável resposta.

O ajuste de parâmetros do modelo foi conduzido com combinações fatoriais de três variáveis: 1) algoritmo de treinamento da rede neural, 2) número de neurônios na camada oculta da rede, e 3) o parâmetro de atraso de tempo no modelo autorregressivo.

Foram realizados experimentos com três algoritmos de treinamento de redes neurais:

1. *trainlm*: função de treinamento de rede que atualiza os valores de peso e viés de acordo com a otimização de Levenberg-Marquardt.
2. *trainbr*: função de treinamento de rede que atualiza os valores de peso e viés de acordo com a otimização de Levenberg-Marquardt. Minimiza uma combinação de erros quadrados e pesos e, em seguida, determina a combinação correta para produzir uma rede que generalize bem. Este processo é chamado de regularização bayesiana.
3. *traincg*: é uma função de treinamento de rede que atualiza os valores de peso e polarização de acordo com o método de gradiente conjugado escalonado.

Os experimentos consideraram também a variação nos parâmetros de atraso  $d$  do modelo NARX. Resultados recentes mostram que esse parâmetro produz um efeito significativo nos resultados de previsão do modelo e dependem, em grande medida, de características dos dados como, por exemplo, a taxa de amostragem das séries temporais. O atraso determina quantos valores passados serão usados pelos modelos autorregressivos para estimar o valor atual da série temporal. Nos experimentos desse trabalho foram considerados os valores de atrasos iguais a 4, 7, 11 e 14 dias, tendo em conta tanto a frequência diária de amostragem dos dados e as características reais do sistema sendo modelado quanto ao tempo de incubação da doença.

Por fim, também foram testados o número de neurônios na camada interna, com os valores: 10, 15, 20 e 25. Definir a quantidade de neurônios na camada oculta é importante já que um número excessivo de neurônios pode acarretar *overfitting*. De maneira oposta, um número pequeno de neurônios na camada oculta pode não ser suficiente para capturar a relação não linear entre as variáveis dependentes e independentes do modelo (BRAGA et al., 2000). Vale destacar que, nesse caso, os valores experimentados foram escolhidos tanto com base na literatura quanto por produzirem os melhores resultados em experimentos iniciais. Existem técnicas que avaliam o efeito da quantidade de neurônio na camada oculta, como se analisa em Sa'ad et al. (2022), mas que não foram estudadas com detalhes e utilizadas neste trabalho.

A combinação fatorial das variáveis gerou uma sequência de 48 séries de testes. Para cada série foram executadas 30 replicações seguindo uma metodologia baseada no teorema do limite central, a fim de quantificar a incerteza associada com a utilização da técnica na estimativa para que os erros encontrados no processo de calibração do modelo estejam em conformidade com uma distribuição de probabilidade normal (VUOLO, 1996).

Os resultados dos experimentos executados foram avaliados de forma a encontrar as combinações dos fatores que produzem o menor erro, aferido por meio do erro quadrático médio (MSE, do inglês *Mean Squared Error*).

Com base nos resultados obtidos nesse primeiro conjunto de experimentos, foi definida uma segunda etapa de testes, variando o número de camadas ocultas da rede e a função de ativação dos neurônios nessas camadas, bem como mudando a transformação dos dados de entrada da normalização para padronização.

### 3.5 CENÁRIOS DE DISTANCIAMENTO SOCIAL

O distanciamento social teve importante papel na redução dos casos de COVID-19 e consequente na redução de óbitos, porém efeitos colaterais em diversos setores da sociedade foram inevitáveis. Como exemplo, na pesquisa de Levandowski et.al. (2021), ficou evidenciado que o distanciamento social causou uma queda no número de notificações de violência contra crianças no Brasil, exigindo que fossem criadas estratégias para identificação dos casos suspeitos durante a pandemia. Os

autores afirmam, que a subnotificação interfere em medidas de planejamento e ações nas áreas de saúde, proteção social, justiça e segurança pública.

Estes efeitos colaterais do distanciamento social também foram pesquisados nas áreas de saúde. Pachioni et al (2020), analisa em seu trabalho o aumento de diagnósticos de depressão durante a pandemia. Estimativas apontam que houve aumento de 25% nos casos no primeiro ano de pandemia, o que desperta a preocupação dos governantes sobre como administrar as consequências de longo prazo da COVID-19. As medidas foram implementadas num cenário emergencial, com preocupação exclusiva de não esgotamento de leitos de hospitais e reduções do número de mortos.

Desta forma, observa-se a necessidade de analisar o distanciamento social para entendimento do seu impacto e avaliar se medidas mais ou menos restritivas teriam resultados diferentes. Esse trabalho, então, manipula as séries de índice de distanciamento social em patamares menores e maiores do que os efetivamente observados na prática.

Os valores de distanciamento social para os cenários simulados,  $c_1(t)$ , em relação aos valores efetivamente observados,  $x(t)$ , são apresentados na Tabela 3. Os intervalos foram definidos para garantir que, estatisticamente, os cenários fossem significativamente diferentes.

Tabela 3: Cenários hipotéticos de distanciamento social considerando os valores de distanciamento efetivamente observados nos últimos  $N$  dias de observação, sendo  $N$  o horizonte de previsão ( $N=7$  ou  $N=14$ ).

Cenário ( $q$ )	Distanciamento social	Expressão
#1	75% menor	$c_1(t) = 0,25 \cdot x(t)$
#2	50% menor	$c_2(t) = 0,50 \cdot x(t)$
#3	25% menor	$c_3(t) = 0,75 \cdot x(t)$
#4	25% maior	$c_4(t) = 1,25 \cdot x(t)$
#5	50% maior	$c_5(t) = 1,50 \cdot x(t)$

Fonte: Autor

Os dados da série temporal de distanciamento social são decrescidos e acrescidos em escalas de 25%, conforme Tabela 3. Essas simulações são realizadas com o objetivo de entender como a redução e o aumento do distanciamento social afetaria o número de novos casos e óbitos.



Os cenários são avaliados por uma versão *closed-loop* do modelo a qual realiza uma previsão múltiplos passos à frente a partir dos dados simulados de distanciamento social. Nesse caso, valores preditos também são usados como entrada para o modelo para estimar o próximo passo de tempo conforme horizonte de previsão estipulado,  $N$ . A formulação matemática que expressa essa relação é ilustrada na equação (4), como um exemplo para  $N = 5$  e  $d = 7$ , na qual  $\widetilde{y}_q(t)$  denota o valor estimado da variável de interesse (novos casos ou óbitos) para o cenário  $q$ ,  $q = 1, \dots, 5$ . Importante notar na equação (4) que o valor de  $\widetilde{y}_q$  representa o número esperado de novos casos (ou óbitos) quando a variável exógena é representada pelos índices de distanciamento social definidos na Tabela 7, e denotados por  $c_q$ ,  $q = 1, \dots, 5$ .

Além disso, é importante destacar que os valores simulados de distanciamento social são utilizados apenas para as previsões futuras. Isso significa que tanto no treinamento do modelo quanto na previsão de  $y(t+1)$  os valores reais de distanciamento social são usados. Simula-se, portanto, uma modificação no índice de distanciamento a partir do tempo  $t+1$  apenas.

$$\begin{aligned}
 \widetilde{y}_q(t+1) &= F(y(t), y(t-1), \dots, y(t-6), x(t), x(t-1), \dots, x(t-6)) \\
 \widetilde{y}_q(t+2) &= F(\widetilde{y}_q(t+1), y(t), y(t-1), \dots, y(t-5), c_q(t+1), x(t), \dots, x(t-5)) \\
 \widetilde{y}_q(t+3) &= F(\widetilde{y}_q(t+2), \widetilde{y}_q(t+1), \dots, y(t-4), c_q(t+2), c_q(t+1), \dots, x(t-4)) \\
 \widetilde{y}_q(t+4) &= F(\widetilde{y}_q(t+3), \widetilde{y}_q(t+2), \dots, y(t-3), c_q(t+3), c_q(t+2), \dots, x(t-3)) \\
 \widetilde{y}_q(t+5) &= F(\widetilde{y}_q(t+4), \widetilde{y}_q(t+3), \dots, y(t-2), c_q(t+4), c_q(t+3), \dots, x(t-2))
 \end{aligned} \tag{4}$$

No capítulo 4 são descritos os resultados de calibragem dos parâmetros do modelo e os resultados da simulação dos cenários de distanciamento social. Foram realizadas simulações na rede NARX para cada faixa de distanciamento social, de acordo com as Tabelas 4, 5, 6 e 7, as quais consideram os melhores valores dos parâmetros observados na primeira fase de testes.

Tabela 4: Parâmetros utilizados nos experimentos com diferentes cenários para predição de novos casos no horizonte de previsão de 7 dias.

Cenário	Simulação do distanciamento social	Atraso	Algoritmo	Neurônios	Horizonte de previsão
Cenário 1	75% menor	11	trainlm	10	7 dias
Cenário 2	50% menor	11	trainlm	10	7 dias
Cenário 3	25% menor	11	trainlm	10	7 dias
Cenário 4	25% maior	11	trainlm	10	7 dias
Cenário 5	50% maior	11	trainlm	10	7 dias

Fonte: Autor

Tabela 5: Parâmetros utilizados nos experimentos com diferentes cenários para predição de novos casos no horizonte de previsão de 14 dias

Cenário	Simulação do distanciamento social	Atraso	Algoritmo	Neurônios	Horizonte de previsão
Cenário 1	75% menor	4	trainbr	15	14 dias
Cenário 2	50% menor	4	trainbr	15	14 dias
Cenário 3	25% menor	4	trainbr	15	14 dias
Cenário 4	25% maior	4	trainbr	15	14 dias
Cenário 5	50% maior	4	trainbr	15	14 dias

Fonte: Autor

Tabela 6: Testes realizados com diferentes cenários para predição de óbitos no horizonte de previsão de 7 dias

Cenário	Simulação do distanciamento social	Atraso	Algoritmo	Neurônios	Horizonte de previsão
Cenário 1	75% menor	14	trainbr	15	7 dias
Cenário 2	50% menor	14	trainbr	15	7 dias
Cenário 3	25% menor	14	trainbr	15	7 dias
Cenário 4	25% maior	14	trainbr	15	7 dias
Cenário 5	50% maior	14	trainbr	15	7 dias

Fonte: Autor

Tabela 7: Testes realizados com diferentes cenários para predição de óbitos no horizonte de previsão de 14 dias

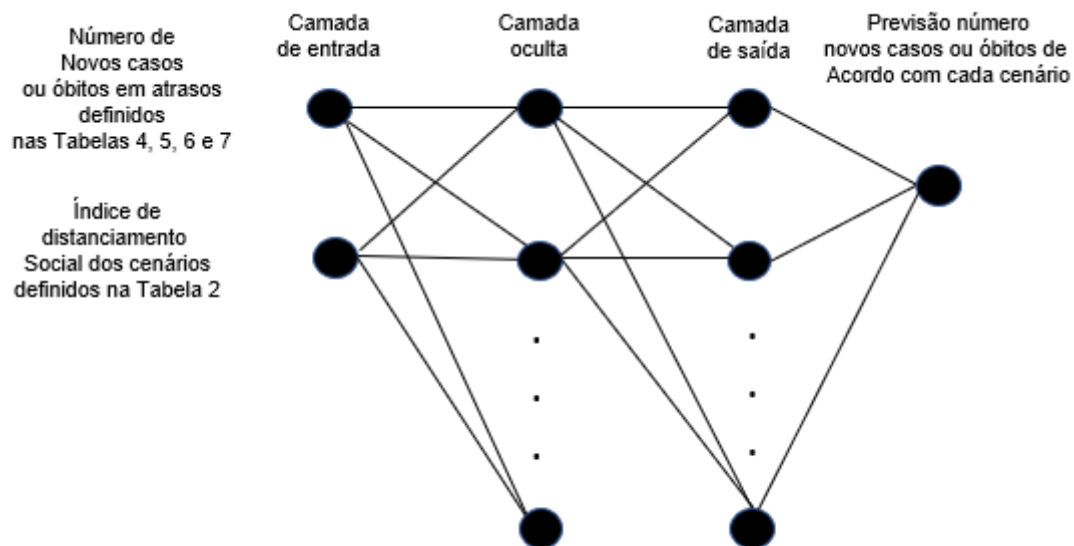
Cenário	Simulação do distanciamento social	Atraso	Algoritmo	Neurônios	Horizonte de previsão
Cenário 1	75% menor	14	trainbr	20	14 dias
Cenário 2	50% menor	14	trainbr	20	14 dias
Cenário 3	25% menor	14	trainbr	20	14 dias
Cenário 4	25% maior	14	trainbr	20	14 dias
Cenário 5	50% maior	14	trainbr	20	14 dias

Fonte: Autor

A Figura 9 mostra o esquema de redes neurais para simulação dos cenários de distanciamento social. Os dados de novos casos ou óbitos ( $y$ ) são as entradas da rede NARX em atrasos de tempo definidos nas Tabelas 4, 5, 6 e 7. Assim os valores de

entrada são:  $y(t-1)$ ,  $y(t-2)$ ,  $y(t-3)$ , ...,  $y(t-d)$ , onde  $d$  é o atraso encontrado que obtém os melhores resultados do MSE na primeira etapa dos experimentos.

Figura 9: Esquema de rede neural para os cenários de distanciamento social



Fonte: Autor

### 3.7 FUNÇÃO DE ATIVAÇÃO

Todos os experimentos foram também realizados com as funções de ativação tangente sigmoide (tansig) e sigmoide logística (logsig), porém não se observou melhores estimativas do MSE. Assim, todos os resultados apresentados foram obtidos com a função de ativação tangente sigmoide.

## 4 RESULTADOS E DISCUSSÃO

O planejamento fatorial completo possibilitou a análise de todas as combinações dos parâmetros de ajustes da rede neural. O objetivo é o ajuste de parâmetros do modelo NARX, no qual a variável externa é o índice de distanciamento social.

Na tabela 8 são apresentados os resultados de MSE sobre a série temporal normalizada e padronizada.

Tabela 8: Melhores resultados obtidos nos experimentos computacionais com planejamento fatorial completo.

$N$	$d$	$y$	Número neurônios	algoritmo	MSE (normalização)	MSE (padronização)
7	11	Novos casos	10	trainlm	0,00280	1,009e-06
7	14	Óbitos	15	trainbr	0,00560	0,002500
14	4	Novos casos	15	trainbr	0,01820	5,987e-07
14	14	Óbitos	20	trainbr	0,01950	0,004200

Fonte: Autor

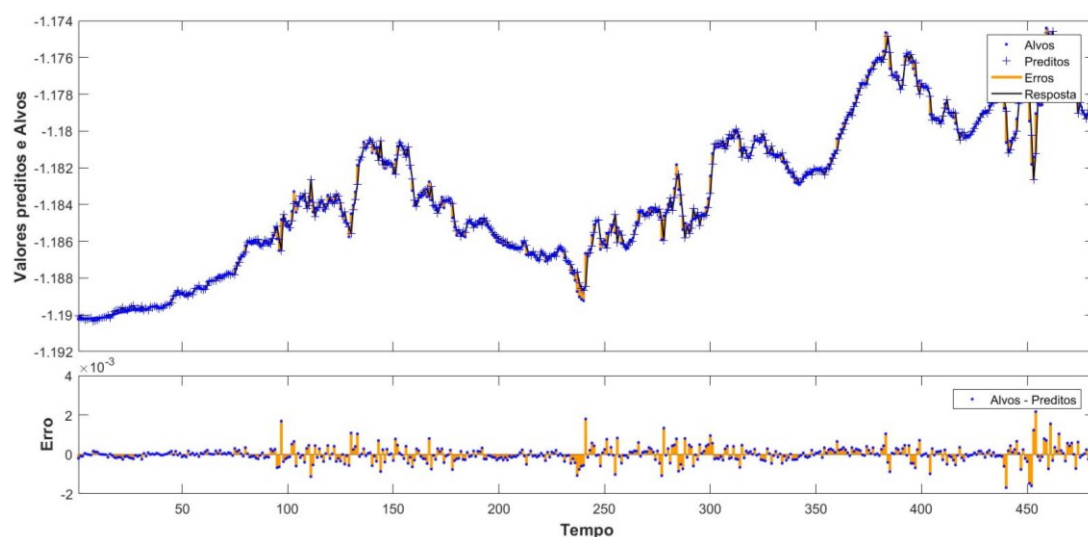
A Tabela 8 mostra que a padronização das séries temporais possibilitou melhores resultados que a normalização. Entretanto, as previsões para o número de óbitos careciam de melhorias. Estes dados permitem uma discussão sobre as melhores técnicas estatísticas para preparação dos dados para treinamento de redes neurais. A padronização possibilitou melhores resultados de MSE, indicando que, para a aplicação tratada neste trabalho, esse método é melhor que a normalização para o pré-processamento dos dados de treinamento da rede neural tipo NARX. Uma possível explicação para esse resultado é o fato dessa transformação ser menos sensível à presença de *outliers*, como é o caso da base de dados utilizada. Portanto, todos os resultados apresentados a seguir são referentes à aplicação da transformação de padronização aos dados.

Adicionalmente, é possível observar que existe uma relação mais forte entre o índice de distanciamento social e o número de novos casos em comparação ao número de óbitos, o que é um resultado esperado, uma vez que o óbito por COVID-19 está ligado a uma série de outros fatores que vão muito além do distanciamento como, por exemplo, a idade (maior que 60 anos) e a presença de comorbidades (doenças cardiovasculares, obesidade, diabetes, hipertensão, doenças pulmonares

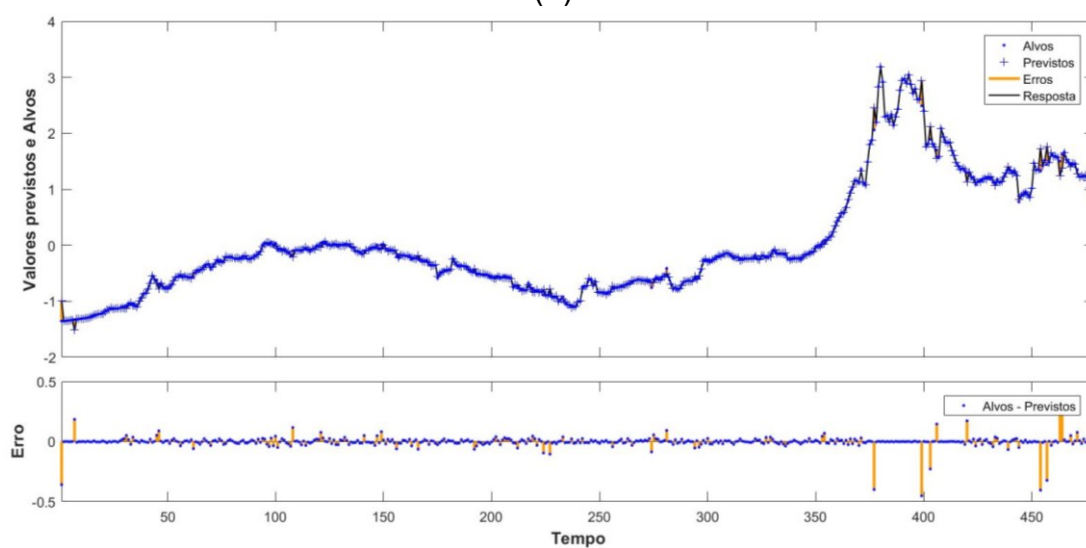
etc.) (KOLIFARHOOD et al., 2020). Mesmo o uso de um número maior de neurônios na camada oculta e um maior atraso não foi suficiente para atingir um valor de MSE equivalente ao obtido para variável número de novos casos. Especificamente, com um aumento no número de neurônios de 10 para 15, com horizonte de previsão  $N=7$ , e de 15 para 20 com horizonte  $N=14$ , e um aumento no atraso no modelo autorregressivo (de 11 para 14 com  $N=7$  e de 4 para 14 com  $N=14$ ), o resultado para o número de óbitos foi consistentemente pior. Esse resultado indica que o comportamento da variável número de óbitos é mais difícil de ser estimado quando se considera apenas o distanciamento social como variável exógena. Ainda assim, os resultados de previsão do modelo são promissores, tanto em relação ao índice de distanciamento social e número de novos casos da doença.

Um exemplo do desempenho do treinamento dos modelos é apresentado pela Figura 9, que ilustra uma comparação entre valores alvos e preditos pelo modelo acompanhados dos respectivos erros de previsão, para o caso  $N=7$  e ambas as variáveis de interesse. Observa-se que os erros de previsão para o número de óbitos são relativamente maiores, apesar de apresentarem uma menor variação como pode ser observado nas Figura 10 e 11.

Figura 10: Comparação entre valores alvo e preditos pelo modelo com erros de previsão. Exemplo da primeira iteração para os seguintes exemplos: (a) número de novos casos, modelo com atraso de tempo  $d=11$  dias e 10 neurônios na camada oculta, e (b) variável número de óbitos, atraso  $d=14$  dias e 15 neurônios.



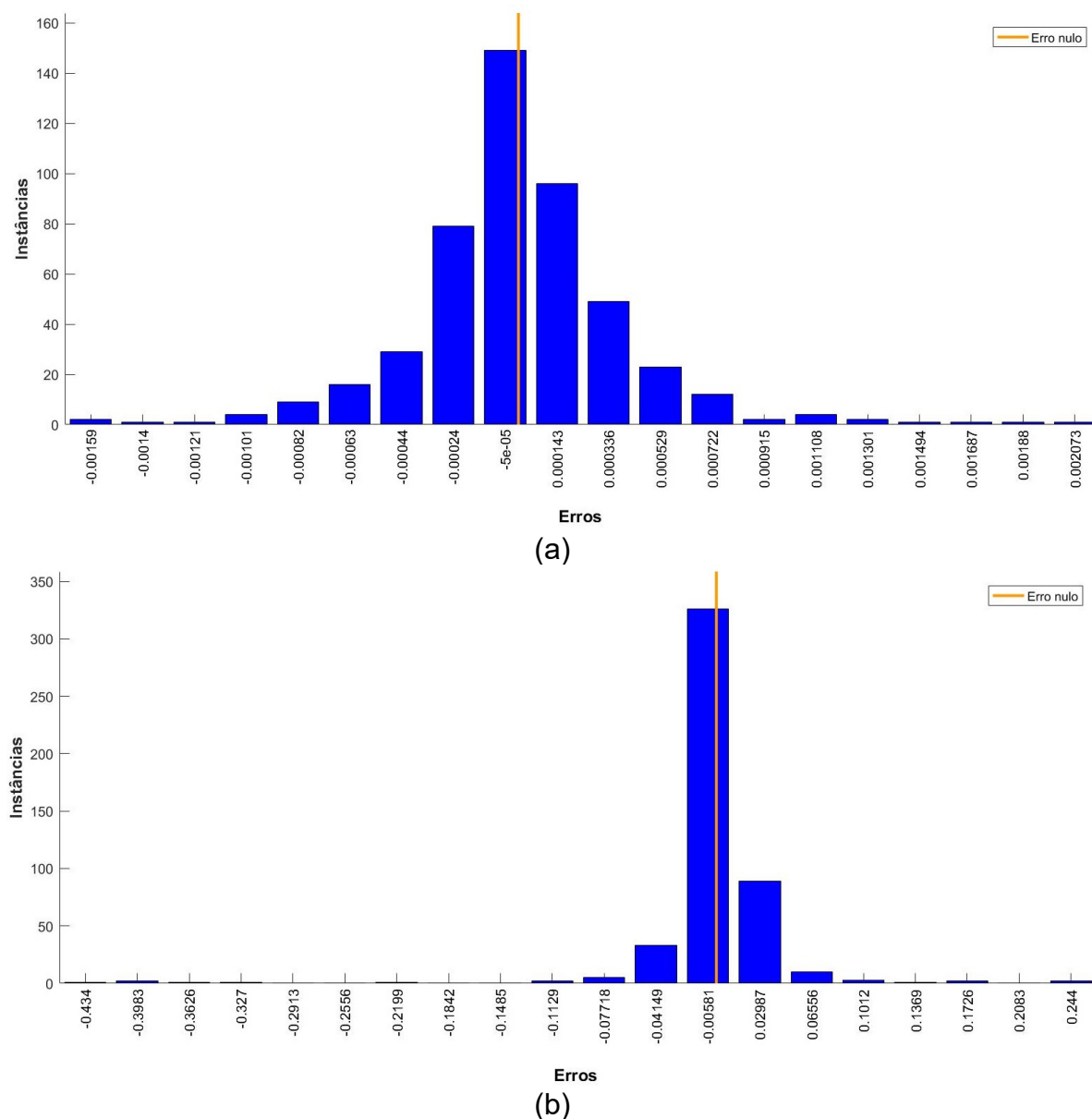
(a)



(b)

Fonte: Autor

Figura 11: Histograma dos erros de previsão dos modelos. Exemplo para a primeira replicação (de um total de 30) com horizonte de previsão  $N=7$ , para variável número de novos casos (a) e número de óbitos (b) para os mesmos exemplos na Figura 9.



Fonte: Autor

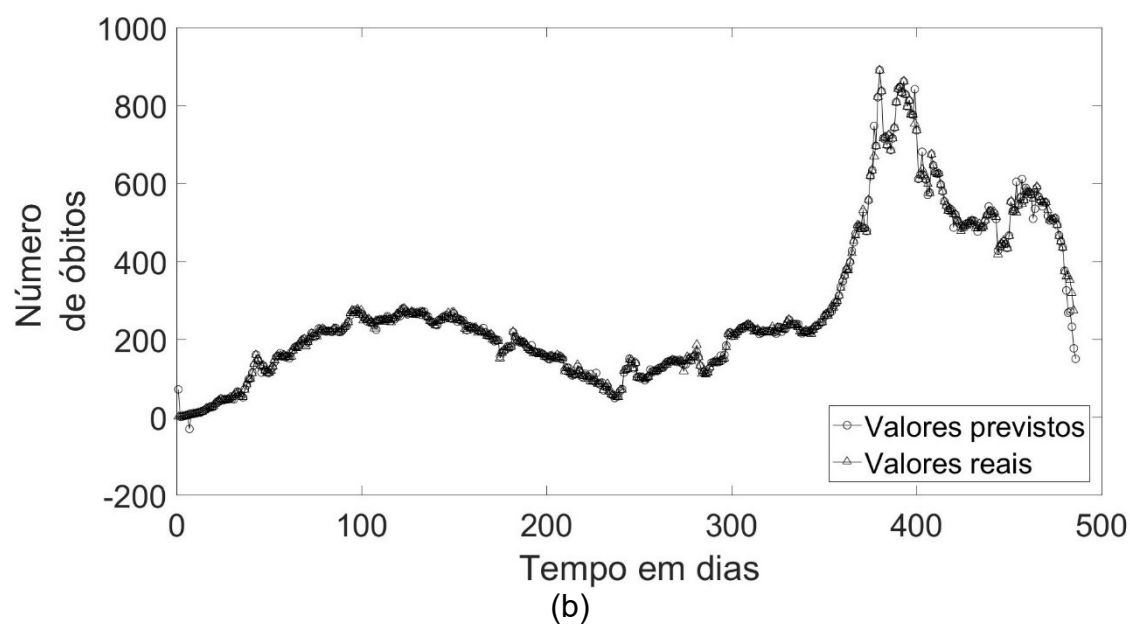
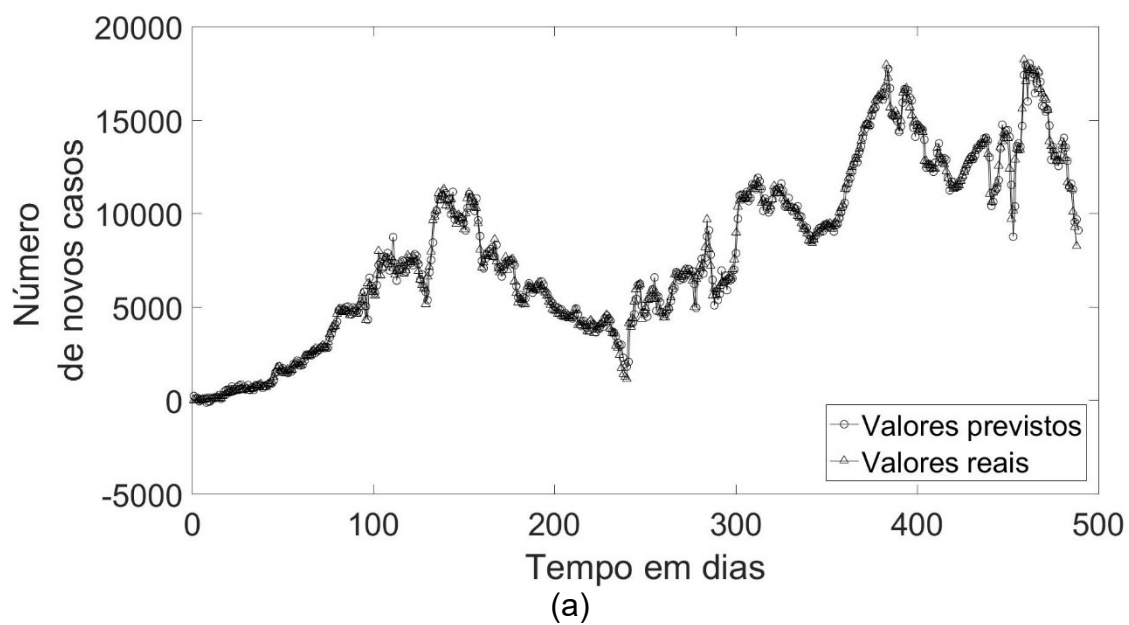
Os modelos treinados foram posteriormente avaliados em sua capacidade de realizar previsões um passo à frente,  $t+1$ , bem como previsões múltiplos passos à frente. Nesse último caso, utiliza-se uma malha fechada na qual o valor predito no passo  $t+1$  é usado como entrada para estimação da variável no instante  $t+2$ , recursivamente.

Os resultados da previsão um passo à frente foram obtidos com a malha aberta, ou seja, a saída ou resposta não influencia a entrada (realimentação), estes são apresentados nas Figuras 12 e 13, para horizontes de previsão  $N=7$  e  $N=14$  respectivamente. Em cada uma das figuras são apresentados resultados para as duas variáveis de interesse, sempre utilizando os parâmetros definidos nas Tabelas 4-7. Convém destacar que a previsão dos últimos passos de tempo ( $N=7$  e  $N=14$ ) é realizada baseada apenas em dados que não foram utilizados no treinamento. Nesse caso, os últimos  $N$  passos de tempo dos dados originais são guardados para comparação dos resultados e para alimentar o modelo com dados reais da variável exógena distanciamento social.

Uma análise mais detalhada nos resultados das Figuras 12 e 13 permite observar que, em geral, houve uma tendência dos modelos em superestimar os valores reais, em especial para o número de óbitos com  $N=14$ , nos últimos passos de tempo, conforme ilustram as Figuras 14 e 15. Entretanto, nota-se que os modelos conseguem capturar as tendências dessas duas variáveis, o que para os objetivos deste trabalho é considerado o mais importante. Vale lembrar que o principal objetivo não é prever os valores exatos das variáveis, mas sim avaliar a relação entre essas variáveis e o distanciamento social para o qual a identificação do comportamento da tendência parece ser suficiente.

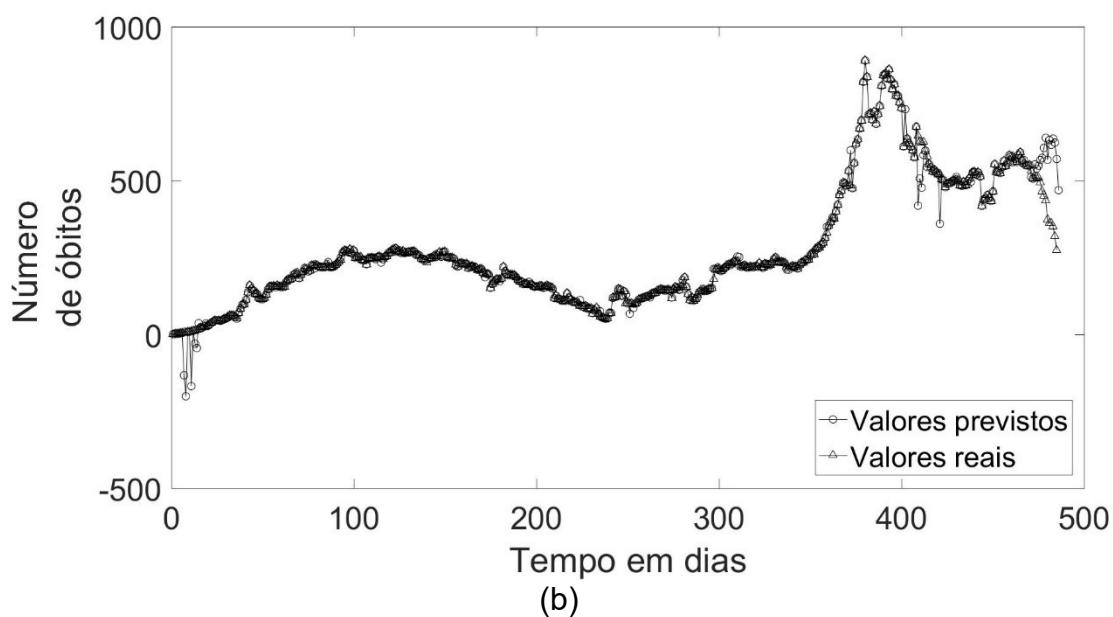
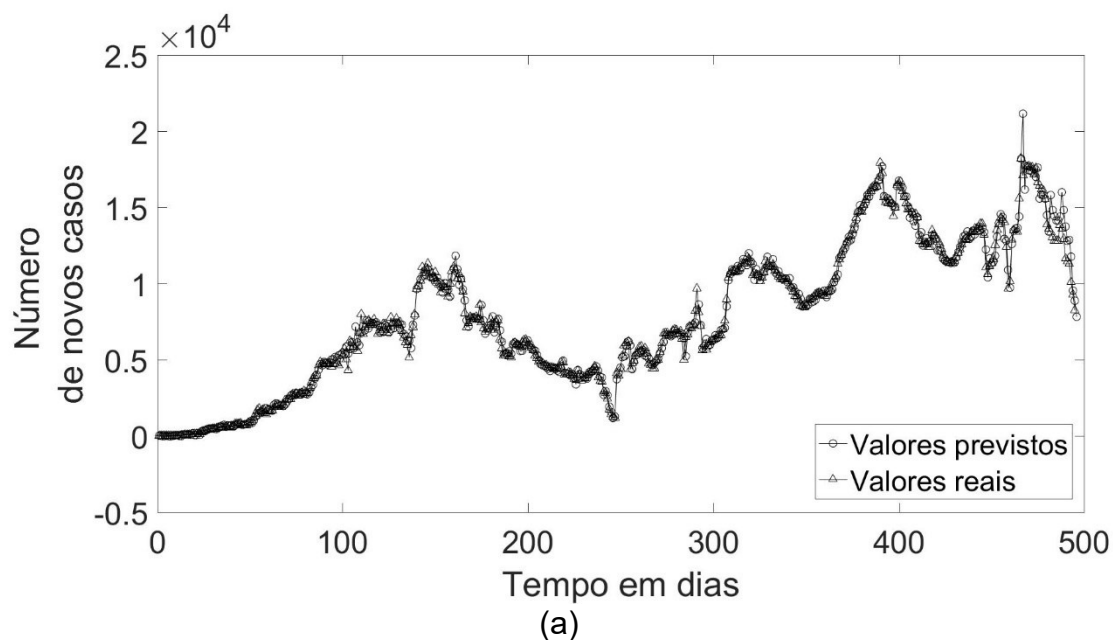


Figura 12: Resultados de previsão com malha aberta para horizonte de previsão  $N=7$ , para o número de novos casos (a) e número de óbitos (b).



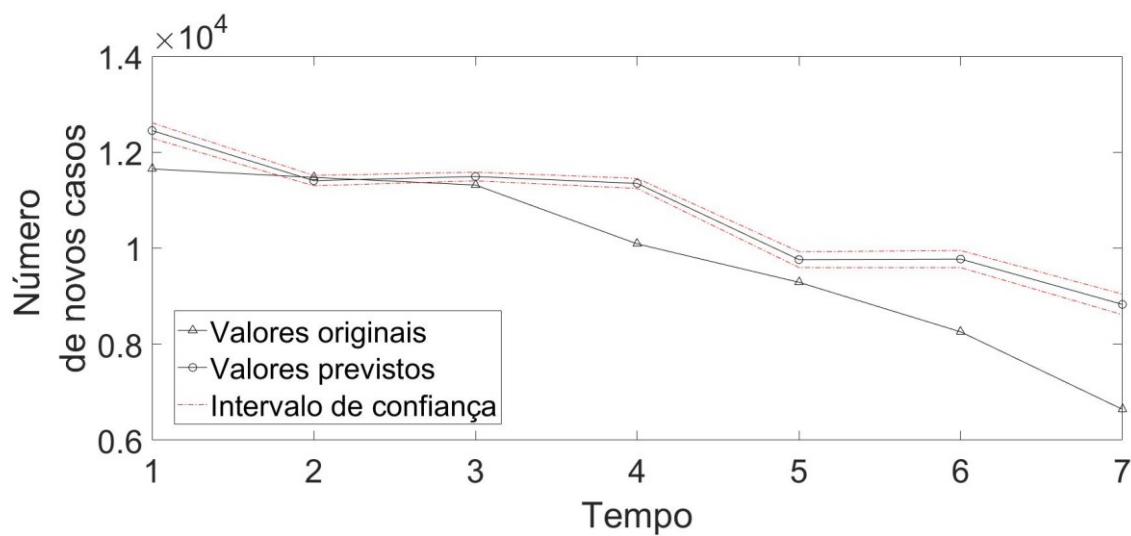
Fonte: Autor

Figura 13: Resultados de previsão com malha aberta para horizonte de previsão  $N=14$ , para o número de novos casos (a) e número de óbitos (b).

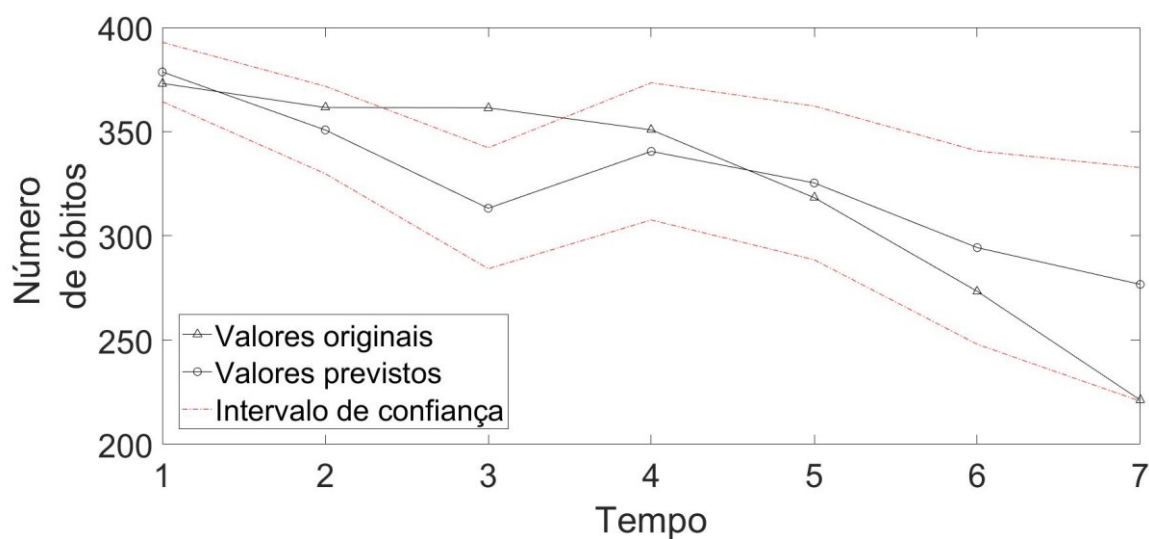


Fonte: Autor

Figura 14: Resultados de previsão com malha aberta para os últimos  $N=7$  passos de tempo, para o número de novos casos (a) e número de óbitos (b).



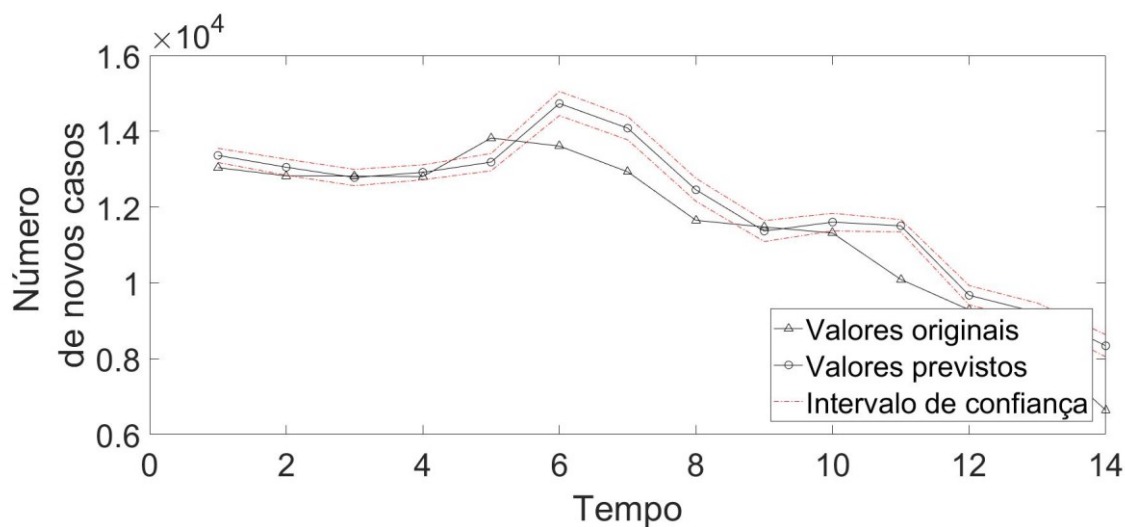
(a)



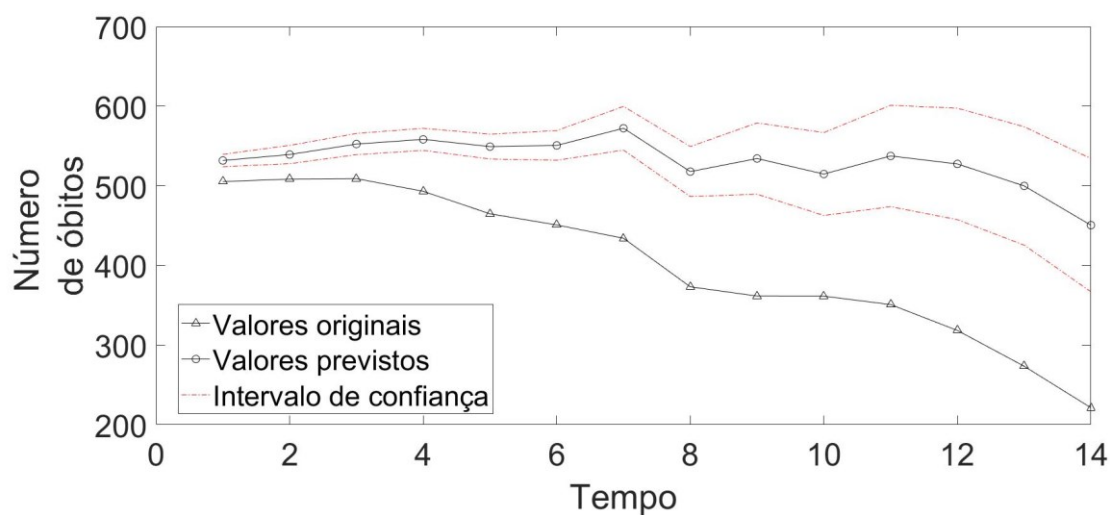
(b)

Fonte: Autor

Figura 15: Resultados de previsão com malha aberta para os últimos  $N=14$  passos de tempo, para o número de novos casos (a) e número de óbitos (b).



(a)



(b)

Fonte: Autor

Importante destacar que os resultados apresentados nas Figuras 14 e 15 se referem à estimativa média das 30 replicações do modelo em cada caso. Foram gerados 30 modelos para cada combinação dos parâmetros definidos nas Tabelas 4-7, o que possibilitou avaliar e quantificar as incertezas nas previsões por meio do cálculo do erro amostral obtido com os resultados das 30 replicações. Os resultados de limite inferior e superior do intervalo de confiança 99% são também ilustrados nas figuras, acompanhados dos valores reais observados em cada variável.

Os limites de confiança apresentados nas Figuras 14 e 15 são uma medida de incerteza associada às previsões feitas por um modelo de rede neural. Quando o intervalo de confiança para as previsões contém os valores originais, pode-se considerar que estatisticamente não há diferença significativa entre os resultados do modelo e os valores reais observados. Esse resultado evidencia a capacidade do modelo em reproduzir a tendência da variável original em todos os casos, ainda que apresente uma tendência de superestimar os valores das séries temporais, mais notadamente para a variável número de óbitos.

É importante destacar que o modelo precisa de menos neurônios na camada interna para prever novos casos quando comparados aos óbitos. Este resultado sugere que o modelo, ao tratar os dados de novos casos generaliza melhor quando comparado aos óbitos. O algoritmo *trainbr*, que utiliza a otimização de *Levenberg-Marquardt*, respondeu como sendo o melhor algoritmo de treino para 3 dos 4 objetos de testes envolvidos nesta pesquisa. No início dos experimentos, através de referências na literatura, o *trainbr* havia se mostrado como aquele que responde melhor a este tipo de problema, no entanto, o algoritmo *trainlm* obteve o melhor MSE para previsão no horizonte de 7 dias para o número de novos casos de COVID-19.

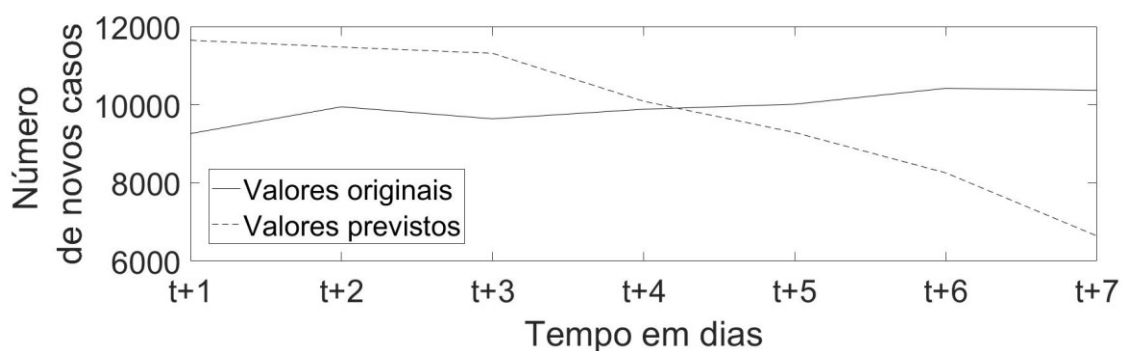
De forma geral, os números apontam que as previsões de novos casos são mais precisas que as de óbitos. A influência do distanciamento social no número de óbitos pode ocorrer em longo prazo, já que outros fatores podem influenciar como: tempo de incubação, demora na resposta das medidas de prevenção e principalmente a capacidade do sistema de saúde, que deve estar preparado para proporcionar atendimento adequado a todos os pacientes. Isso sugere que pode ser necessário considerar mais dias de atraso no modelo para obter melhores resultados na estimativa do número de óbitos.

Em geral, os resultados fornecem indícios sobre a adequação do modelo NARX na assimilação dos padrões nos dados de novos casos de COVID-19 e óbitos, além de serem úteis para indicar quais parâmetros podem contribuir para ajustar melhor o modelo. Vale lembrar, no entanto, que os resultados das Figuras 14 e 15 foram todos obtidos quando os modelos foram usados em malha aberta, utilizando apenas dados reais como entrada. Apesar de isso ser suficiente para representar a relação entre as variáveis, e avaliar efeitos de possíveis variações no índice de distanciamento social, uma análise de cenários futuros requer a previsão de múltiplos à frente o que

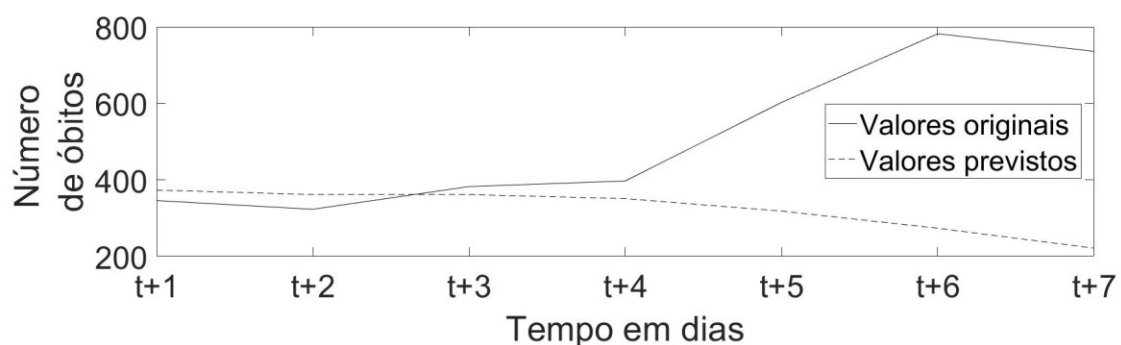
exige dos modelos a capacidade de realizar estimativas precisas também nesse contexto.

Portanto, convém observar os resultados de previsão múltiplos passos à frente em cada um dos modelos criados. Nesse caso, os modelos são usados em uma versão com malha fechada o que significa que os valores preditos em um passo de tempo são imediatamente usados para estimar o valor no passo de tempo seguinte. Pressupõe-se que os valores da variável exógena são conhecidos, sendo usados os  $N$  últimos originais. Os resultados correspondentes são apresentados na Figura 16, seguindo a mesma sequência adotada nas figuras anteriores.

Figura 16: Resultados de previsão  $N=7$  passos à frente para (a) número de novos casos e (b) número de óbitos. Modelos usados em malha fechada.



(a)



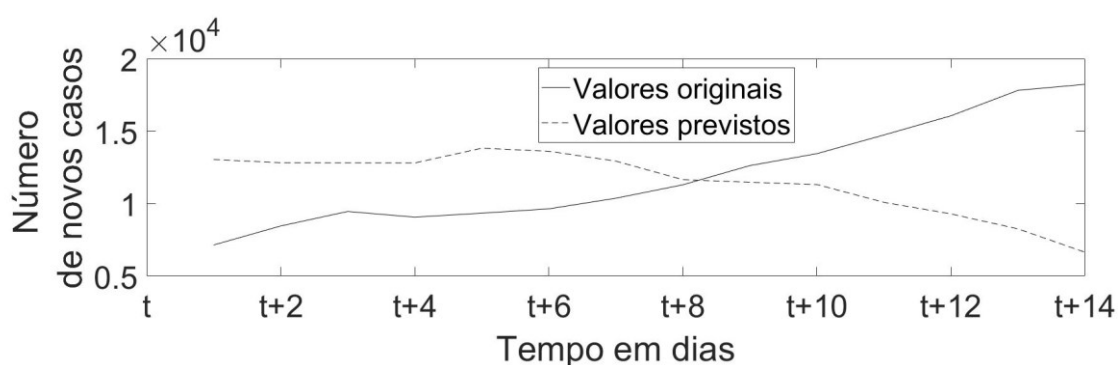
(b)

Fonte: Autor

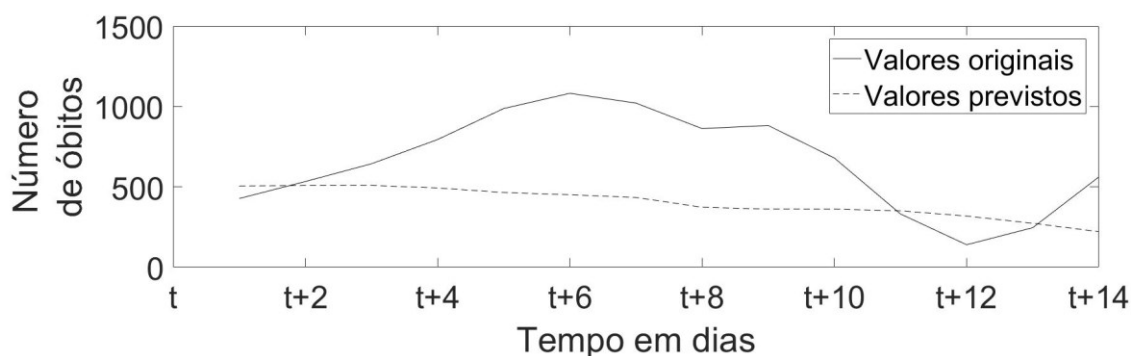
Como pode ser observado na Figura 16, os resultados de previsão múltiplos passos à frente apresentaram um erro médio de cerca de 20% para ambas as variáveis. Exceto para os quatro primeiros passos de previsão do número óbitos, no caso  $N=7$ , todos os demais valores apresentaram uma significativa diferença em

relação ao valor real. Essa diferença foi ainda maior para o horizonte de previsão  $N=14$ , como esperado uma vez que a previsão múltiplos passos à frente é feita com o ciclo fechado em que os valores previstos são usados como entrada para novas previsões, conforme observa-se na Figura 17. Para o exemplo no qual o atraso de tempo do modelo  $d=4$ , previsões a partir do tempo  $t+5$  usam apenas valores preditos como entradas do modelo.

Figura 17: Resultados de previsão  $N=14$  passos à frente para (a) número de novos casos e (b) número de óbitos. Modelos usados em malha fechada.



(a)



(b)

Fonte: Autor

Portanto, antes de seguir para a análise dos cenários de distanciamento social definidos na Tabela 3, foi necessário aprimorar a capacidade dos modelos de realizar previsões múltiplos passos à frente. Considerando os resultados do planejamento de experimentos inicial e avaliando o impacto das variações dos parâmetros no desempenho do modelo, optou-se por realizar um novo conjunto de testes com diferentes algoritmos de treinamento no modelo NARX, mantendo os demais

parâmetros inalterados (salvo se explicitamente descrito). Assim, foram realizados testes para os parâmetros definidos na Tabela 8 para cada um dos algoritmos de treinamento descritos resumidamente na Tabela 9. Neste caso, apenas a variável número de novos casos, com  $N=7$ , foi considerada nesta etapa. Como uma análise desses métodos está além do escopo deste trabalho, são indicadas na tabela algumas referências apropriadas que fornecem descrições mais detalhadas e os fundamentos computacionais de cada algoritmo utilizado.

Os resultados para previsão múltiplos passos à frente obtidos com modelos treinados pelos diferentes algoritmos descritos na Tabela 9 são apresentados na Figura 18, considerando um horizonte de previsão  $N=7$  para número de novos casos. Previsões para horizontes de tempo maiores que 7, em malha fechada, bem como para variável número de óbitos, resultaram precisões consideradas insuficientes e não foram consideradas.

A partir da Figura 18 é possível observar que, de fato, o algoritmo de otimização exerceu uma significativa influência nos resultados, o que possivelmente está relacionado à capacidade de generalização do modelo resultante. Além disso, é possível observar que alguns algoritmos de treinamento resultaram em modelos que, apesar de não prever com exatidão o valor real do número de novos casos, foram capazes de capturar a tendência da variável o que, como já mencionado, está coerente com o objetivo proposto. Esse é o caso, por exemplo, dos algoritmos de treinamento *Scaled Conjugate Gradient* (SCG), *Resilient Backpropagation* (RP), *Gradient Descent* (GD) e *Variable Learning Rate Gradient Descent* (GDX).



Tabela 9: Algoritmos de treinamento utilizados nesta etapa do trabalho.

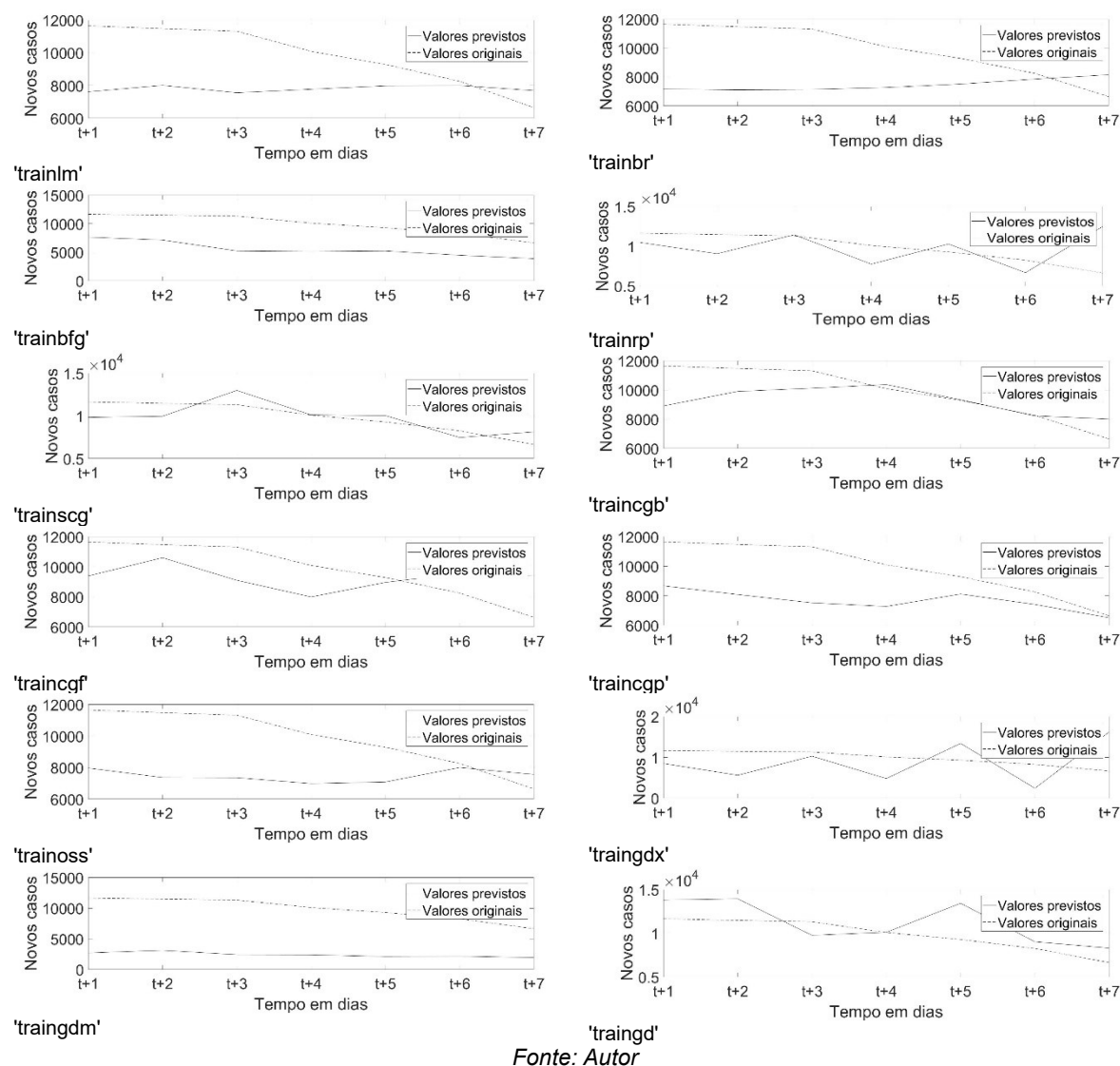
Algoritmo	Notação no Matlab	Descrição e fonte para consulta
Levenberg-Marquardt	'trainlm'	Utiliza a matriz Hessiana para realizar estimativas melhores tanto para tamanhos de passo quanto para direções. (YU e WILARMOWSKI, 2012)
Bayesian Regularization	'trainbr'	Converte uma regressão não linear em um problema estatístico "bem formulado" (DEMUTH e BEALE, 2000).
BFGS Quasi-Newton	'trainbfg'	Realiza uma aproximação do inverso da matriz Hessiana e é usado para resolver minimização quadrática estática (LEITHEAD e ZHANG, 2007)
Resilient Backpropagation	'trainrp'	Utiliza o sinal do gradiente para indicar a direção do ajuste dos pesos. (RIEDMILLER e BRAUN, 1992).
Scaled Conjugate Gradient	'trainscg'	Classe de métodos com o objetivo de minimizar funções suaves (NOCEDAR e WRIGHT, 1999)
Conjugate Gradient with Powell/Beale Restarts	'traincgb'	Atualiza os valores dos pesos e vieses de acordo com o algoritmo de retropropagação conjugado de gradiente com reinícios de Powell-Beale (NIHAN e TULAY, 2003).
Fletcher-Powell Conjugate Gradient	'traincgf'	Atualiza os valores dos pesos e vieses de acordo com o gradiente conjugado de backpropagation. (NIHAN e TULAY, 2003).
Polak-Ribière Conjugate Gradient	'traincgp'	Atualiza os valores de peso e viés de acordo com o método de hackpropagation com atualizações de Polak-Ribiere. (NIHAN e TULAY, 2003).
One Step Secant	'trainoss'	É uma tentativa de preencher a lacuna entre os algoritmos de gradiente conjugado e os algoritmos de Quasi-Newton (NIHAN e TULAY, 2003).
Variable Learning Rate Gradient Descent	'traingdx'	Combina taxa de aprendizado adaptativa com treinamento de momentum, como um parâmetro adicional de treinamento (NIHAN e TULAY, 2003).
Gradient Descent with Momentum	'traingdm'	É um método de um único passo em que a próxima iteração depende apenas do gradiente atual. (POLYAK, 1964).
Gradient Descent	'traingd'	Os pesos e vieses são atualizados na direção do gradiente negativo da função de desempenho (NIHAN e TULAY, 2003).

Fonte: Autor

Os resultados também indicam que existe espaço para uma melhoria adicional nos resultados por meio da variação dos demais parâmetros em conjunto com o algoritmo de treinamento.

A partir dos resultados obtidos com a variação dos algoritmos de treinamento, apresentados na Figura 18, e considerando a variabilidade desses resultados, foram selecionados os melhores modelos com os algoritmos de treinamento SCG, RP, GD e GDX, os quais foram usados para avaliar o impacto de variações no índice de distanciamento social no número de novos casos.

Figura 18: Previsão múltiplos passos à frente, com malha fechada, com diferentes algoritmos de treinamento. Resultados de  $N=7$  para o número de novos casos. Valor de atraso fixado em  $d=11$ .



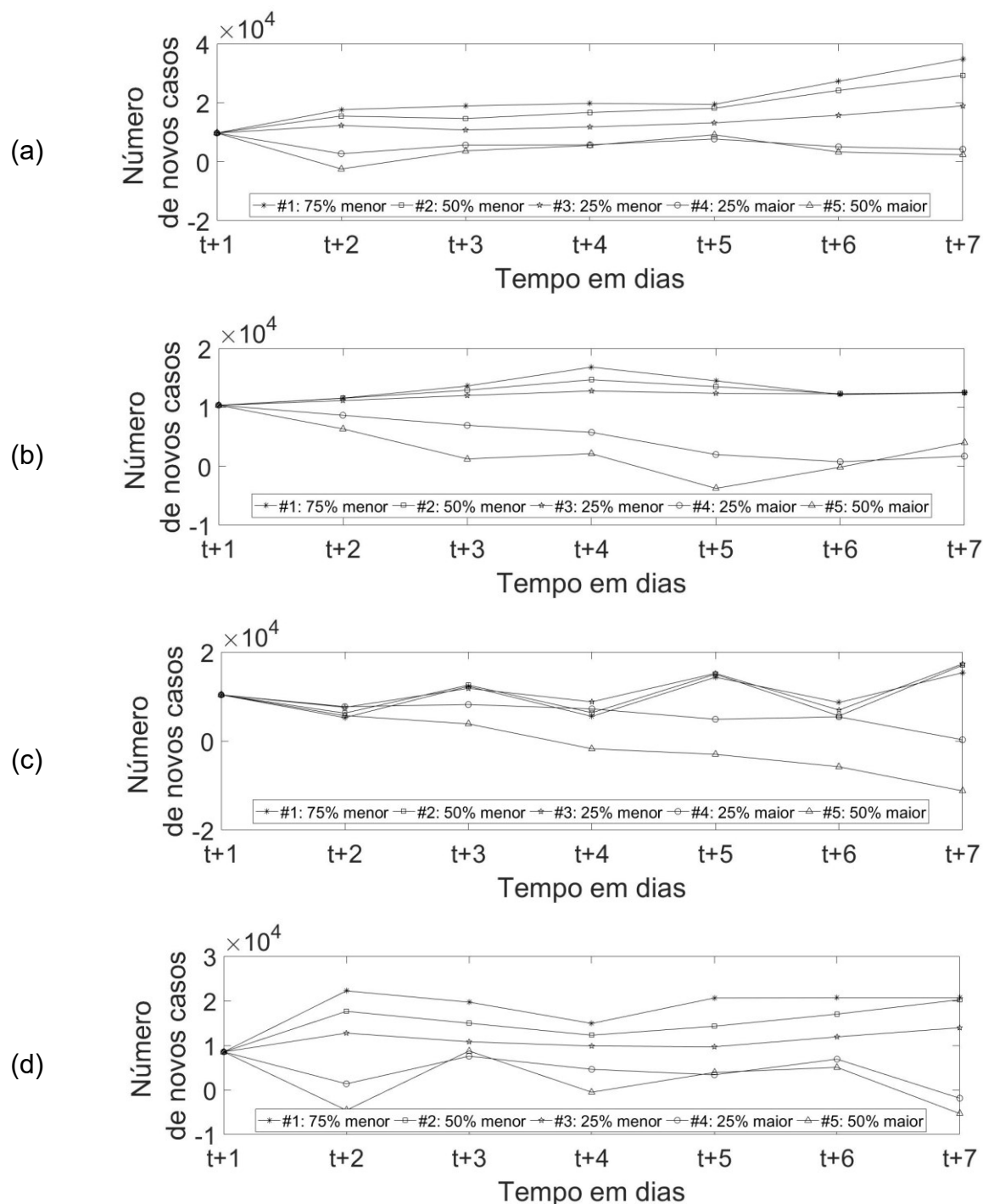
#### 4.1 RESULTADOS DO IMPACTO DO DISTANCIAMENTO SOCIAL

Foram realizados experimentos com cenários hipotéticos de distanciamento social, conforme pode ser observado na Figura 19 para a variável número de novos casos. Considerando o desempenho dos modelos criados em relação à acurácia das estimativas em malha fechada, apenas o horizonte de previsão  $N=7$  foi avaliado nessa etapa.

Apesar das incertezas e variabilidade das estimativas para os modelos, observa-se um comportamento que se repete nas diferentes configurações. Em geral, uma redução no distanciamento social aumenta o número de novos casos, o que é esperado, mas parece haver um limite para o índice de distanciamento a partir do qual o impacto no número de novos casos deixa de ser significativo. Vale destacar, no entanto, que cenários de maior distanciamento produzem significativas reduções tanto no número de casos quanto no número de mortes, o que sugere a eficiência de medidas mais restritivas de distanciamento.

Os resultados desta pesquisa mostram que uma rede NARX pode ser aplicada de forma a assimilar dados de novos casos, óbitos e distanciamento social em um modelo não-linear, gerando assim, um modelo capaz de mensurar o impacto do distanciamento social na pandemia de COVID-19.

Figura 19: Simulação do impacto de diferentes cenários de distanciamento social no número de novos casos. Previsão para  $N=7$  dias, com algoritmo de treinamento (a) SCG, (b) RP, (c) GD e (d) GDX.



Fonte: Autor

Como o objetivo do trabalho é avaliar e quantificar o impacto das medidas de distanciamento social, foi pesquisado um índice que pudesse aferi-lo. O único encontrado, até então, é obtido das redes de telefonia celular, através de um método de distanciamento de um ponto base. Ainda que existam inúmeros trabalhos para

identificação de aglomerações por meio do tratamento de imagens, observa-se que estes dados têm carácter mais focal e não reproduziriam de uma forma mais ampla a adesão à política de distanciamento social e por isso não foram eleitos para uma extrapolação. As imagens de câmeras podem identificar uma aglomeração através do reconhecimento de padrões, como pesquisou Dias et al. (2021), mas não seria possível a generalização para uma cidade, por exemplo.

Os testes foram realizados numa versão de malha fechada (*closed loop*) e os dados previstos são utilizados como entradas para a previsão dos próximos passos. Foram apresentados na Figura 19 simulações para o horizonte de previsão de 7 dias, pois as estimativas para o horizonte de 14 dias se mostraram inadequadas e muito pouco confiáveis. A rede neural NARX apresenta muitas vantagens para modelar relações não lineares, no entanto, há um aumento de complexidade quando se compara às redes neurais de alimentação direta (*feedforward*).

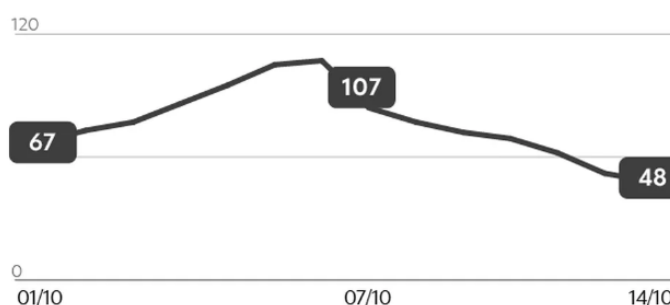
Para se obter um panorama completo de novos casos e óbitos no horizonte de previsão de 7 dias, são necessários, de acordo com os experimentos, 14 dias de histórico. No Brasil, o Consórcio de veículos de imprensa, a partir de dados das secretarias estaduais de saúde, avaliou os casos de COVID-19 sempre comparando com o número de novos casos e óbitos do 14º dia anterior, conforme se ilustrado na Figura 20. Desta forma, considera-se que são necessários, ao mínimo, 14 dias para avaliação de medidas de contenção na pandemia, o que sugere que um atraso máximo 14 dias no modelo NARX são suficientes para o entendimento do impacto das medidas de distanciamento social. Maiores valores de atraso não foram testados neste trabalho, mas poderiam resultar em melhores previsões para o número de óbitos.

Figura 20: Infográfico comparando os números de novos casos com o 14º dia anterior.

### Brasil, 14 de outubro

- **Total de mortes:** 687.186
- **Registro de mortes em 24 horas:** 66
- **Média de mortes nos últimos 7 dias:** 48 (variação em 14 dias: -28%)
- **Total de casos conhecidos confirmados:** 34.788.500
- **Registro de casos conhecidos confirmados em 24 horas:** 6.318
- **Média de novos casos nos últimos 7 dias:** 4.463 (variação em 14 dias: -34%)

### Média móvel de mortes



Fonte: Extraído de <https://g1.globo.com/saude/coronavirus/noticia/2022/10/14/media-movel-de-mortes-por-covid-no-brasil-volta-a-apontar-queda-e-indica-menos-de-50-vitimas-por-dia.ghtml> pelo próprio autor em 9/6/23

Os achados deste trabalho não puderam ser corroborados com dados de outras pesquisas, pois não foram encontradas muitas referências para modelagem não linear dos dados da pandemia. Nota-se predominância de artigos que estudaram análise de imagens para identificar, por exemplo, aglomerações em locais públicos (DIAS et al. 2021; TANWAR et. al. 2021; NIU et al. 2021. Assim, a área da saúde pode se beneficiar de modelos de previsão baseados em inteligência artificial, visto que a aplicação desse conceito na área ainda é relativamente recente. Necessita-se, porém, avaliar a capacidade de generalização desses modelos no que diz respeito à possibilidade de que decisões inteligentes em uma localidade ou uma doença específica, possam ser reproduzidas para outras regiões e/ou outras doenças. Desta forma, estudos como este contribuem para complementar o conhecimento da área.

Destaca-se a importância da escolha da metodologia de modelagem de dados para evitar a possibilidade de gerar uma interpretação equivocada, ainda mais quando a coleta dos dados não segue padrões rígidos. Neste trabalho, a média móvel foi aplicada para promover uma suavização dos dados e minimizar variações,

possivelmente causadas por problemas de subnotificação e acúmulo de casos de COVID-19 em determinados dias da semana.

O trabalho mostra a importância da estatística na preparação dos dados para redes neurais, onde deve-se observar processos de limpeza de dados, estratégias para trabalhar com dados ausentes e detecção de outliers. A normalização e a padronização são processos que colocam os dados em uma mesma escala, com o objetivo de impedir que o treinamento tenha influência de uma determinada variável. O trabalho desperta a atenção para identificação dos melhores modelos estatísticos a serem aplicados nas séries temporais.

Por fim, vale frisar a abordagem proposta pode ser igualmente aplicada em outros tipos de dados epidemiológicos, seja avaliando o efeito do distanciamento social no número de casos ou óbitos por outras doenças ou ainda avaliando o efeito de outras variáveis exógenas como, por exemplo, temperatura e volume de chuvas, ou variáveis ligadas ao aspecto sociodemográfico de uma população.

## 5 CONCLUSÕES

A implementação de medidas públicas durante a pandemia de COVID-19 encontrou diversos desafios, como medir a efetividade das ações de forma isolada, pois o cenário exigia diversas intervenções ao mesmo tempo. Neste sentido, Ferguson et al. (2020) citam que as medidas teriam pouca efetividade se tivessem sido implementadas de maneira isolada. Uma destas medidas é o distanciamento social, que, neste trabalho, é quantificado por meio de dados de mobilidade. Foram encontradas duas formas de quantificar o distanciamento social: (1) dados de mobilidade do Google e (2) dados de antenas de celulares do SIMI. A opção pelos dados do SIMI-SP possibilitou uma visão mais ampla do cenário e foi considerada em todas as etapas do trabalho. Observa-se que a literatura trouxe poucas referências sobre esse assunto.

A revisão da literatura ainda apontou que a maioria dos trabalhos relacionados ao distanciamento social teve como objetivo a identificação de aglomerações, através do reconhecimento de padrões em imagens. Nota-se que muitos destes estudos foram derivados de trabalhos anteriores de identificação de padrões. Vale frisar que o surgimento da pandemia também motivou muitas pesquisas sobre modelos que ajudassem a interpretar e fazer projeções no cenário de COVID-19. Nesse contexto, porém, predominaram adaptações e implementações do modelo SIR, os quais definem os possíveis estados de um indivíduo durante uma epidemia. Conforme apresentado nas delimitações, o trabalho não se preocupa com a dinâmica da pandemia e sim, em oferecer estimativas que possam contribuir para entendimento da relação entre novos casos e óbitos com o distanciamento social. Assim, foram utilizadas redes neurais artificiais em um modelo onde fosse possível obter estimativas que sugeressem a efetividade das medidas de distanciamento social. O trabalho não questiona nem critica as políticas de distanciamento social, tendo apenas como objetivo quantificar o impacto do mesmo na pandemia de COVID-19.

A pesquisa também mostrou a possibilidade de utilização de redes neurais em detrimento a modelos puramente estatísticos. Conforme observado na revisão bibliográfica, as aplicações de NARX, na maioria dos casos, tem relação com previsões de demandas, manutenção de sistemas mecânicos e previsões financeiras. Adicionalmente, destaca-se a possibilidade de ampliar o entendimento da modelagem



NARX aplicadas a um modelo autorregressivo no contexto da epidemiologia matemática.

Através dos experimentos, pode-se concluir que a fase de preparação dos dados para as redes neurais é uma tarefa muito importante. A normalização e a padronização são técnicas utilizadas comumente na fase de preparação dos dados, e tem o objetivo de colocar os dados em um intervalo de valores comuns. As mesmas, como se observou nos experimentos, podem impactar na acurácia e podem interferir na forma de avaliar o modelo. No trabalho, a padronização ajudou a função de ativação tangente sigmoide a resolver melhor o problema de regressão.

É importante destacar que a acurácia de uma rede neural depende dos dados de treinamento, sendo assim, problemas de dados faltantes e subnotificações comprometem a confiança nas estimativas geradas e precisam ser considerados para uso destas informações para ações de planejamento do sistema de saúde. Para fins de análise da série temporal, não foi possível verificar a série de dados do índice do distanciamento social nos dias atuais, pois a captação se encerrou em 31/12/21. Esta análise possibilitaria avaliar a confiança nos dados por meio da constatação de índices de distanciamento social possivelmente mais baixos, já que quase não existem medidas restritivas de circulação de pessoas em 2023.

A literatura especializada indica que medidas de restrição de mobilidade da população geram efeitos significativos na vida de indivíduos e que os efeitos colaterais do distanciamento social precisam ser considerados. O entendimento da quantificação do impacto do distanciamento social não apenas pode auxiliar em estratégias de contenção da proliferação de vírus, mas em medidas de mitigação de possíveis efeitos negativos.

Os cenários hipotéticos, nos quais o índice de distanciamento social é manipulado em patamares maiores ou menores que os reais índices, estimam que o distanciamento social influencia o número de novos casos e óbitos, numa relação inversamente proporcional, mas que parece haver um limiar para o índice de distanciamento a partir do qual essa influência perde significado. Tais estimativas podem colaborar em pesquisas futuras para o desenvolvimento de ferramentas que forneçam informações para que autoridades possam tomar decisões antecipadamente e garantindo assim, a definição de medidas para sistemas de saúde.

De forma geral, observando a revisão bibliográfica, os trabalhos sobre redes neurais pesquisam: experimentos de novas arquiteturas, aplicações em diferentes

áreas de conhecimento, comparativo de diferentes algoritmos de treinamento, otimização e análise de interpretabilidade. Baseado nisso, os resultados deste trabalho são apresentados na Tabela 10.

Tabela 10: Painel teórico sobre pesquisas de redes neurais comparando com os achados desse trabalho

<b>Objetivo</b>	<b>Achados neste trabalho</b>
Experimentos de novas arquiteturas	O trabalho não traz nenhuma contribuição a respeito
Aplicações em diferentes áreas de conhecimento	A pesquisa mostra a possibilidade de aplicações da rede NARX na epidemiologia matemática
Comparativo de diferentes algoritmos de treinamento	Os experimentos testaram, por meio de um planejamento fatorial completo, diferentes algoritmos de treinamento
Otimização	Os resultados foram otimizados a partir da implementação da normalização e padronização
Análise de interpretabilidade	O trabalho interpreta os resultados com base na influência do distanciamento social no número de novos casos e óbitos e não entrega somente estimativas.

*Fonte: Autor*

O aumento da produção científica durante a pandemia foi considerável em algumas áreas da saúde como epidemiologia, virologia e medicina. A urgência motivou cientistas a buscarem respostas para o problema, já que a algum tempo o mundo não passava por uma situação como o COVID-19. A quantidade de informação gerada foi muito grande e abre-se a discussão sobre o compartilhamento rápido e fácil de dados. O Brasil é um país diverso, onde as regiões têm características muito distintas em relação ao clima, acesso ao sistema de saúde e principalmente sobre questões sociais, desta forma destaca-se a necessidade de ações colaborativas entre institutos de pesquisa e universidades sobre a análise de dados de COVID-19 em larga escala, de forma que seja possível uma visão mais generalista sobre o assunto.

## 5.1 SUGESTÕES DE TRABALHOS FUTUROS

A execução dos experimentos mostrou oportunidades para serem pesquisadas e seguem como proposta de continuidade deste trabalho.

O modelo NARX proporciona uma realimentação dos dados de entrada a partir das saídas da rede, como a previsão de múltiplos passos à frente. No entanto, somente a série de interesse, neste caso, o número de novos casos e mortos é realimentada. Uma proposta seria o desenvolvimento de outra rede complementar para previsão dos dados de distanciamento social.

É necessário também uma pesquisa para avaliar a possibilidade de generalização do modelo para outras regiões e doenças. A variável exógena utilizada neste trabalho poderia representar outras séries temporais, como temperatura e fatores sociais, de forma que o modelo pudesse trazer diferentes formas de interpretação.

Durante o desenvolvimento do trabalho foi observado que as técnicas de preparação dos dados impactam na acurácia dos resultados, desta forma, um estudo detalhado sobre normalização e padronização de dados para redes neurais pode ser estudado com o objetivo de comparar o resultado em diferentes aplicações. Ainda na etapa de pré-processamento, pesquisas complementares com outras técnicas de suavização de dados podem ser realizadas, neste trabalho foi utilizada a média móvel de 7 dias, já que esta foi definida como parâmetro de monitoramento da evolução da pandemia de COVID-19, porém outras técnicas podem ser aplicadas, como a transformada wavelet, por exemplo.

## Referências bibliográficas

ADITYA SATRIO, B. et al. Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET. **Procedia Computer Science**, 2021. 524-532.

ALEMNEH, H. T.; KASSA, A. S.; GODANA, A. A. An optimal control model with cost effectiveness analysis of Maize streak virus disease in maize plant. **Infectious Disease Modelling**, 2021. 169-182.

ANDERSON, R. M. et al. How will country-based mitigation measures influence the course of the COVID-19 epidemic ? **The Lancet**, 2020.

ANDRADE, Maria Margarida de. **Introdução à metodologia do trabalho científico**. 10 ed. São Paulo: Atlas, 2017.

AQUINO, E.; SILVEIRA, I. H.; PESCARINI, J. M. Medidas de distanciamento social no controle da pandemia de COVID-19: potenciais impactos e desafios no Brasil. **Ciência e Saúde Coletiva**, 2020.

ARAUJO, R. et al. A deep increasing-decreasing-linear neural network for financial time series prediction. **Neurocomputing**, 347, 2019. 59-81.

ARAZI, R.; FEIGEL, A. Discontinuous transitions of social distancing in the SIR model. **Physica A: Statistical Mechanics and its Applications**, 566, Março 2021. 125632.

B. T. Polyak, "Some methods of speeding up the convergence of iteration methods," **USSR Comput. Math. Math. Phys.**, vol. 4, no. 5, pp. 1–17, 1964

BOLDOG P, TEKELI T, Vizi Z, Denes A, Bartha FA, Rost G. Risk Assessment of Novel Coronavirus COVID-19 Outbreaks Outside China. **J Clin Med** 2020; 9:2.

BRASIL, Ministério da Saúde. Gerência de Epidemiologia de campo. **Boletim Epidemiológico**, Brasília, Maio/22, Disponível em: <

[https://www.saude.df.gov.br/documents/37103/0/Boletim\\_Covid\\_746.pdf](https://www.saude.df.gov.br/documents/37103/0/Boletim_Covid_746.pdf)>. Acessado em 19/06/2023.

BONAT, W. H.; RIBEIRO Jr, P. J.; ZEVIANI, W. M. Regression models with responses on the unity interval: specification, estimation and comparison. **Biometrical Brazilian Journal**, v. 30, p. 415–431, 2012.

CHATERJEE, S. et al. Software fault prediction using Nonlinear Autorregressive with exogenous inputs (NARX) network. **Applied Intelligence**, v. 37, n.1, p.121-129, 2012

CHIMMULA, V. K. R.; , L. Z. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. **Chaos, Solitons & Fractals**, 135, Junho 2020.

CHINAZZI M, Davis JT, AJELLI M, GIOANNINI C, LITVINOVA M, MERLER S, PIONTTI AP, MU K, ROSSI L, SUN K, VIBOUD C. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. **Science** 2020; pii:eaba9757

COCCIA, M. Improving preparedness for next pandemics: Max level of COVID-19 vaccinations without social impositions to design effective health policy and avoid flawed democracies, **Environmental Research**, 213, 2022, 113566.

COVID-19: Relatório de mobilidade da comunidade. Google. Disponível em: <<https://www.google.com/covid19/mobility/>>. Acesso em: 24 de abril de 2023.

DEMUTH, H.; BEALE, M. Neural Network Toolbox User's Guide Version 4; **The Math Works Inc.**: Natick, MA, USA, 2000; pp. 5–22.

DEZOTTI, C. ; YANG, H. M. Teoria do Raio Espectral e uma Caracterização do número de reprodutibilidade basal para doenças de transmissão direta. **Biomatemática**, 1999.

DIAS, Bruno S. S., et al. Sistema para a identificação de aglomerações operando em Redes IoT e Fog Computing: **Revista Ibérica de sistemas e tecnologias de informação**, 25 agosto de 2021

FEDELE, R.; MERENDA, M. An IoT System for Social Distancing and Emergency Management in Smart Cities Using Multi-Sensor Data. **Algorithms**, 29 Setembro 2020.

FELDMANN, R. M.; LINHARES GARCIA, E.; KRUG, S. B. Consumo de drogas em tempos de pandemia: reflexões sobre impactos na saúde mental. **Interthesis**, 30 Maio 2022.

FERGUSON, N. M.; LAYDON, D.; NEDJATI-GILANI, G. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. **Imperial College COVID-19 Response Team**, 16 Março 2020.

GALVÃO, C.O.; VALENÇA, M.J.S. (org.) Sistemas Inteligentes: aplicações a recursos hídricos a ciências ambientais. Porto Alegre: **Editora da Universidade**, UFRGS, ABRH, 1999. 246 P.

GAO, G. F. From “a” to “z”: attacks from emerging and re-emerging pathogens. **PubMed**, 2018.

GARCEL, A.; SOUZA NETO, J. L. Do "Stay home" ao "Lockdown". O Impacto das medidas de distanciamento social no Brasil e no mundo. **Revista de Teorias da Democracia e Direitos Políticos**, 22 Dezembro 2020.

GENSINI, GF, YACOUUB , MH e CONTI, AA. The concept of quarantine in history: from plague to SARS. **J Infect** 2004;49(4):257-61.

GREENSTONE, M.; NIGAM, V. Does Social Distancing matter ? **SSRN Product & Service**, 27 Março 2020.

GRIETTE, Q.; MAGAL, P. Clarifying predictions for COVID-19 from testing data: The example of New York State. **Infect Dis Model**, 13 Janeiro 2021.

HOU, C.; CHEN, J.; ZHOU, Y. The effectiveness of quarantine of Wuhan city against the Corona Virus Disease 2019 (COVID-19): A well-mixed SEIR model analysis. **Journal of Medical Virology**, 25 Abril 2020.

HOUVÊSSOU GM, SOUZA TP, Silveira MF. Medidas de contenção de tipo lockdown para prevenção e controle da COVID-19: estudo ecológico descritivo, com dados da África do Sul, Alemanha, Brasil, Espanha, Estados Unidos, Itália e Nova Zelândia, fevereiro – agosto de 2020. **Epidemiol. SerV. Saúde** [preprint]. 2020 [citado 2020 nov 12]:[21 p.]. Disponível em: <https://doi.org/10.1590/s1679-49742021000100023>

HUANG, L.; QIN, J.; ZHOU, Y. LIU, L., SHAO L. Normalization Techniques in Training DNNs: Methodology, Analysis and Application. **Cornell University**, 27 Setembro de 2020

Instituto Brasileiro de Geografia, Estatística (IBGE). Síntese de indicadores sociais: uma análise das condições de vida da população brasileira. Rio de Janeiro: IBGE; 2018

NOCEDAL, J.; S. WRIGHT S. J., Numerical Optimization, **Springer**, Berlin, 1999.

KALIL, S.J.; MAUGERI, F.; RODRIGUES, M.I. Response surface analysis and simulation as a tool for bioprocess design and optimization. **Process Biochemistry**, v.35, p.539-550, 2000

KERMACK, W.O.; MCKENDRICK, A.G. A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London. **Proceedings of the royal society A**, 01 Agosto 1927. A115: 700-721.

KUCHARSKI AJ, RUSSELL TW, DIAMOND C, LIU Y, EDMUNDS J, FUNK S, EGGO RM, Sun F, JIT M, MUNDAY JD, Davies N. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. **Lancet Infect Dis** 2020; S1473-3099(20)30144-4

KOLIFARHOOD, G. et al. Epidemiological and Clinical Aspects of Covid-19: a narrative review. p. 9, **Pubmed Central** 2020.

KRAEMER MU, YANG CH, GUTIERREZ B, WU CH, KLEIN B, PIGOTT DM, DU PLESSIS L, FARIA NR, LI, R., HANAGE WP, BROWSTEIN JS. The effect of human mobility and control measures on the COVID-19 epidemic in China. **Science** 2020; pii:eabb4218

KRATSCHMER V. Strong consistency of least-squares estimation in linear regression models with vague concepts. **J Multivar Anal** 2006;97(3):633–54.

LAU H., KHOSRAWIPOUR V, KOCBACH P., MIKOLAJCZYK A, SCHUBERT J, BANIA J, KHOSRAWIPOUR T. The positive impact of lockdown in Wuhan on containing the COVID-19 outbreak in China. *J Travel Med* 2020; pii:taaa037.

LEON, C.; ORTEGA, F. Nowcasting Economic Activity with Electronic Payments Data: A Predictive Modeling Approach. **Economia del Rosario**, 18 Junho 2018. 381-407.

LEVANDOWSKI, M.L. Impacto do distanciamento social nas notificações de violência contra crianças e adolescentes no Rio Grande do Sul, Brasil. **Caderno de Saúde Pública**, 6 de novembro de 2020

LI, L. et al. Propagation analysis and prediction of the covid-19. **Infectious Disease Modelling**, v. 5, p. 282 – 292, 2020.

LIESTOL K, ANDERSEN P.K. Updating of covariates and choice of time origin in survival analysis: problems with vaguely defined disease states. **Stat Med** 2002;21(23):3701–14.

LOUNIS, M; TORREALBA-RODRIGUES, O; CONDE-GUTIERREZ, R A. Predictive models for COVID-19 cases, deaths and recoveries in Algeria, 2021

MAHAYANA, D. et al., Stability Analysis of Covid 19 SIR Modeling. **International Journal on Electrical Engineering and Informatics**, Volume 13, Março 2021



MARINOV, T. T.; MARINOVA, R. S. Dynamics of COVID-19 using inverse problem for coefficient identification in SIR epidemic models. **Chaos, Solitons & Fractals: X**, Março 2020.

FREEDMAN, D. Isolation, quarantine, social distancing and community containment: pivotal role for old-style public health measures in the novel coronavirus (2019-nCoV) outbreak. **Journal of travel medicine**, 10 Fevereiro 2020.

MOLLENHIFF, K., BRETZ, F., & DETTE, H. Equivalence of regression curves sharing common parameters, 2019.

MORAES BATISTA, A. F. et al. COVID-19 diagnosis prediction in emergency care patients: a machine learning approach. **The preprint server for health sciences**, 14 Abril 2020.

NETO, B. G. Em tempos de quarentena, uma busca de sua origem. **Enfermagem Brasil**, 06 Junho 2020.

NGUYEN, C.; SAPUTRA, Y. M.; HUYNH, N. V. Enabling and Emerging Technologies for Social distancing: A comprehensive survey and open problems. **Physics and Society**, 22 Setembro 2020.

NICOLAU, L. F. Identificação e previsão neural de atuadores com memória de forma na presença de deformação residual significativa. 2018. PPGEM, UFPB.

NIHAN, C., TULAY, Y. The effects of training algorithms in MLP network on image classification, **IEEE Xplore**, 2003

NISTANE, V. M. Wavelet-based features for prognosis of degradation in rolling element bearing with non-linear autoregressive neural network. **Australian Journal of Mechanical Engineering**, 19, 2019.

NIU, Y. et al. Monocular Pedestrian 3D Localization for Social Distance Monitoring. **Sensors**, 2 Setembro 2021.

NUGHARA, A. et al. Sentiment Analysis on Social Distancing and Physical Distancing on Twitter Social Media using Recurrent Neural Network (RNN) Algorithm. **Journal Online Informatika**, 5, 2020.

OLIVEIRA LATORRE, M. D. R. D.; ALVES CARDOSO, M. R. Análise de séries temporais em epidemiologia: uma introdução sobre os aspectos metodológicos. **Revista Brasileira de Epidemiologia**, Novembro 2001.

PANZAC D. Quarantines and lazarettos: Europe and the plague of the East, XVII–XX centuries [in French]. Aix-en-Provence (France): **Éditions Édisud**; 1986. p. 173–6.

PERAKIS et al. A multiwave SIR-Based model for learning waves, **Productions and operations management**, 2022.

PITOMBEIRA, Delane Felinto; OLIVEIRA, Lucia Conde Pobreza e desigualdades sociais, tensões entre direitos, austeridade e suas implicações na atenção primária. **Ciência e saúde coletiva**, Maio 2020

POKHREL, A. Is MonkeyPox the next pandemic to be dealt with? **Annals of Medicine and Surgery**, 80, 2022, 104131.

QUINTERO, Y. et al. Machine learning models for the prediction of the SEIRD variables for the COVID-19 pandemic based on a deep dependence analysis of variables. **Computers in biology and medicine**, 24 Maio 2021.

REZAEI, M.; AZARMI, M. DeepSOCIAL: Social Distancing Monitoring and Infection Risk Assessment in COVID-19 Pandemic. **Applied Sciences**, 2020.

RIBEIRO, M. H. D. M. et al. Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil. **Chaos, Solitons & Fractals**, v. 135, 2020.

RIEDMILLER, M.; BRAUN, H. 1992. RPROP. A fast adaptive learning algorithm. In Proceedings of the 1992 **International Symposium on Computer and Information Sciences**, Antalya, Turkey, pp.279-285

ROCHA C. PIRES,. Os efeitos sobre grupos sociais e territórios vulnerabilizados das medidas de engrentamento à crise sanitária da COVID-19: Propostas para o aperfeiçoamento da saúde pública. **Instituto de Pesquisa Econômica Aplicada (Ipea)** , 2020.

SA'AD, A.; ZIAD, H.; NYOUNGUE, A. A Comparative Study of Hidden Layer Neurons and Delay Time for Wind energy forecasting model using NARX. 2021 **10th International Conference on Renewable Energy Research and Application (ICRERA)**. [S.l.]: [s.n.]. 2021. p. 47-51.

SAUSEN, A. et al. Classification of the social distance during the COVID-19 pandemic from electricity consumption using artificial intelligence. **International Journal of Energy Research**, 45, 26 Janeiro 2020. 8837-8847.

SESAGIRI, R.; TAN, S. G.; WEE, H. L. Use of Health Belief Model-Based Deep Learning Classifiers for COVID-19 Social Media Content to Examine Public Perceptions of Physical Distancing: Model Development and Case Study. **JMIR Public Health and Surveillance**, 14 Julho 2020.

SCHABENBERGER, O.; PIERCE, F. J. Contemporary Statistical Models for the Plant and Soil Sciences. Boca Raton: **CRC Press**, 2002.

SUPRATID, S.; ARIBARG, T.; SUPHARATID, S. An integration of stationary wavelet transform and nonlinear autoregressive neural network with exogenous input for baseline and future forecasting of reservoir inflow. **Water Resour Manage**, 31, 01 Junho 2017. 4023–4043.

TANWAR, S. et al. Blockchain and AI-Empowered Social Distancing Scheme to Combat COVID-19 Situations. **IEEE Access**, 9, 2021. 129830-129840.

THACKER, P.G.; MENAKER, R.; KOLBE, A.B.; CONNERS, A.L.; AMRAMI, K.K.; CALLSTROM, M.R.; WOOD. C.P. Preparing for the next pandemic: It is more than just about numbers, **Clinical Imaging**, 79, 2021, 179–182.

TRIGUIS SCHIMIT, Pedro Henrique, A model based on cellular automata to estimate the social isolation impact on COVID-19 spreading in Brazi. **Computer Methods and Programs in Biomedicine**, Março 2021.

TSAI, Y.-L. et al. Routing algorithms as tools for integrating social distancing with emergency evacuation. **Scientific Reports**, 4 Outubro 2021.

VUOLO, J. H. Fundamentos da teoria dos erros. 2.ed. São Paulo: Blucher, 1996. 249p.

WANG, Q. et al. Future Trend Forecast by Empirical Wavelet transform and autoregressive moving average, 2018.

W. E. Leithead, Y. Zhang, “ $O(N^2)$ -operation approximation of covariance matrix inverse in Gaussian process regression based on quasiNewton BFGS method,” Commun. Stat. Simul. Comput., vol. 36, no. 2, pp. 367–380, Mar. 2007

WERNECK, G. L.; CARVALHO, M. S. A pandemia de COVID-19 no Brasil: crônica de uma crise sanitária anunciada. **Cadernos de saúde pública**, Rio de Janeiro, Maio 2020.

WOOD, S. N. Inferring UK COVID-19 fatal infection trajectories from daily mortality data: Were infections already in decline before the UK lockdowns? **Biometrics**, 7 Abril 2021.

YANG, D. et al. A Vision-Based Social Distancing and Critical Density Detection System for COVID-19. **Sensors**, 5 Julho 2021.

YU, H.; WILAMOWSKI B.M. Neural Network Training with second order algorithms. **Department of electrical and Computer Engineering**, 2012