DOI: https://doi.org/10.36489/saudecoletiva.2021v11i67p6851-6862

Predição de casos de COVID-19 nos municípios de santa catarina utilizando redes neurais recorrentes

Forecasting COVID-19 cases in santa catarina cities with recurrent neural networks

Predicción de casos de COVID-19 en ciudades de santa catarina utilizando redes neurales recurrentes

RESUMO
Objetivo: objetiva-se avaliar a predição da incidência diária de COVID-19 nos municípios de Santa Catarina, através de um algoritmo de aprendizagem de máquina, em um horizonte de tempo de 14 dias. Método: uma rede neural recorrente foi utilizada para modelar um problema de regressão com propósito preditivo, através de um estudo epidemiológico longitudinal retrospectivo da incidência de COVID-19 nos municípios analisados. Resultados: o modelo de dados obtido através do algoritmo de aprendizagem de máquina apresentou um RMSE de 20,74, menor do que a linha de base estabelecida através de um modelo de persistência. Conclusão: com o resultado alcançado pelo modelo de dados, conclui-se que emprego de ferramentas de Inteligência Artificial permite a obtenção de um importante instrumento para o enfrentamento da pandemia de COVID-19, proporcionando um aperfeiçoamento no gerenciamento de recursos de saúde, que precisam ser adequadamente alocados para uma resposta sanitária adequada ao avanço da doença.
DESCRITORES: Inteligência Artificial; Aprendizado de Máquina; Infecções por Coronavirus; Vigilância em Saúde Pública; Epidemiologia.

ABSTRACT
Objective: evaluate the forecasting of COVID-19 daily incidence in the cities of Santa Catarina, through a machine learning algorithm, within a time horizon of 14 days. Method: a recurrent neural network was applied to model a regression problem with a predictive purpose, using a retrospective longitudinal epidemiological study of COVID-19 cases in the analyzed cities. Results: the data model obtained with a machine learning algorithm presented an RMSE of 20.74, less than the baseline established through a persistence model. Conclusion: from the result achieved by the data model, it follows that the Artificial Intelligence tools used in the research are important instruments to face the COVID-19 pandemic, providing the management improvement of health resources, which need a suitable allocation for an adequate sanitary response to the disease progress.
DESCRIPTORS: Artificial Intelligence; Machine Learning; Coronavirus Infections; Public Health Surveillance; Epidemiology.

RESUMEN
Objetivo: el objetivo es evaluar la predicción de la incidencia diaria de COVID-19 en los municipios de Santa Catarina, mediante la ejecución de un algoritmo de aprendizaje automático, en un horizonte temporal de 14 días. Método: se utilizó una red neuronal recurrente para modelar un problema de regresión con propósito predictivo, mediante un estudio epidemiológico longitudinal retrospectivo de la incidencia de COVID-19 en los municipios analizados. Resultados: el modelo de datos obtenido mediante el algoritmo de aprendizaje automático presentó un RMSE de 20,74, inferior a la línea de base establecida mediante un modelo de persistencia. Conclusión: con el resultado alcanzado por el modelo de datos, se concluye que el uso de herramientas de Inteligencia Artificial permite obtener un instrumento importante para enfrentar la pandemia COVID-19, proporcionando una mejora en la gestión de los recursos de salud, los cuales deben ser adecuadamente asignados para una adecuada respuesta sanitaria al avance de la enfermedad.
DESCRIPTORES: Inteligencia Artificial; Aprendizaje Automático; Infecciones por Coronavirus; Vigilancia en Salud Pública; Epidemiología.

RECEBIDO EM: 02/03/2021 **APROVADO EM:** 10/03/2021

Leonardo Silva Vianna
Odontólogo. Mestre pelo Programa de Pós-Graduação em Informática em Saúde da Universidade Federal de Santa Catarina – PEN/UFSC. Professor do Departamento de Odontologia do Centro Universitário Avantis – UNIAVAN. Balneário Camboriú (SC).
ORCID: 0000-0003-4947-0938

Juliano de Amorim Busana
Enfermeiro. Doutorando do Programa de Pós-Graduação em Enfermagem da Universidade Federal de Santa Catarina – PEN/UFSC. Professor do Departamento de Enfermagem do Centro Universitário Avantis – UNIAVAN. Balneário Camboriú (SC).
ORCID: 0000-0001-7004-2917

INTRODUÇÃO

No decorrer da história da humanidade, foi possível observar a existência de diversas pandemias, caracterizadas através da confirmação de uma rápida disseminação de uma doença em grande parte do globo terrestre. As pandemias ameaçam a saúde de toda a população, obrigando Estado e empresas a se mobilizarem, imbuídos da missão de amenizar as perdas de vidas humanas e também das perdas econômicas[1]. Entre as diversas pandemias avassaladoras que a humanidade enfrentou estão a peste bubônica da Idade Média, a gripe espanhola do século 20 e a gripe suína na última década. Mesmo após a SARS-CoV-1 em 2002 ter assolado o mundo com milhares de mortes, nenhuma vacina foi desenvolvida, nem medicamentos foram criados contra o coronavírus. E conforme descrito por Paumgartten et al.[2], "a pandemia COVID-19 pode ser tudo menos imprevisível".

Apesar da atual pandemia ter se iniciado oficialmente na China em dezembro de 2019, no Brasil, o primeiro caso oficial foi notificado em março de 2020. O delongamento tornou possível a implementação de políticas preventivas pelo Estado Brasileiro, diante da análise do desenvolvimento da doença nos demais países[3]. Entretanto, gestores de saúde, estaduais e municipais, empregaram estratégias díspares, sendo algumas delas alinhadas às recomendações da OMS e de especialistas da área e outras estratégias desalinhadas a tais orientações. Para o adequado enfrentamento da pandemia de COVID-19 faz-se necessário o processamento dos dados existentes sobre a doença, possibilitando sua adequada análise epidemiológica e posterior compartilhamento da informação. Desta forma, os sistemas de Vigilância em Saúde podem ser aprimorados para permitir o acompanhamento em tempo real da doença[4].

A avaliação da dinâmica de transmissão de doenças infecciosas, utilizando modelos matemáticos obtidos pelo processamento de dados, pode subsidiar o planejamento adequado das estratégias de intervenção sanitária. As predições realizadas através dos modelos permitem a alocação adequada de recursos, evitando a redundância na distribuição de profissionais de saúde, bem como podem melhorar a efetividade das campanhas de vacinação[5]. Segundo Bontempi et al.[6], o processamento de dados, executado através de algoritmos de aprendizagem de máquina, permite a predição da ocorrência através da análise dos dados históricos, possibilitando a compreensão de eventos em diferentes áreas do conhecimento. E para que isso seja possível, os algoritmos são empregados na solução de problemas não lineares e não paramétricos, os quais testes estatísticos clássicos não conseguem solucionar. A predição de informações epidemiológicas aprimora as intervenções nos picos de ocorrência das doenças, podendo ser obtida através da aplicação de algoritmos de aprendizagem de máquina para análise de padrões temporais complexos[7].

No decorrer da história da humanidade, foi possível observar a existência de diversas pandemias, caracterizadas através da confirmação de uma rápida disseminação de uma doença em grande parte do globo terrestre.

Desta forma, o objetivo do presente estudo é avaliar a predição da incidência diária de COVID-19 nos municípios de Santa Catarina, através da execução de uma modelagem de dados com um algoritmo de aprendizagem de máquina, em um horizonte de tempo de 14 dias. Determinar uma abordagem adequada para construção de um modelo de dados permite a obtenção de um instrumento que pode aperfeiçoar a capacidade de gestores públicos e administradores de instituições de saúde, para o gerenciamento de recursos de saúde no enfrentamento da pandemia.

MÉTODO

A partir do objetivo da pesquisa, um estudo epidemiológico longitudinal retrospectivo da incidência de COVID-19 nos municípios do Estado de Santa Catarina foi delimitado. O período da disseminação da doença no território, considerado na pesquisa, iniciou em 9 de março de 2020 (data do registro dos primeiros exames com resultado positivo) até 24 de janeiro de 2021. Esse intervalo de tempo foi determinado de modo a ajustar-se a uma periodicidade semanal, totalizando 46 semanas.

Os registros de casos de COVID-19 foram obtidos no repositório de dados mantido pelo Estado de Santa Catarina. Em 26 de janeiro de 2021, o arquivo foi descarregado do Portal de Dados Abertos, que, de acordo com as informações apresentadas no próprio sítio eletrônico, trata-se

de uma plataforma oficial de dados governamentais abertos, que tem o propósito de melhorar a transparência e o controle social, integrando informações do sistema e-SUS Vigilância Epidemiológica e do Sistema de Informação de Vigilância Epidemiológica da Gripe (SIVEP Gripe). Também havia informação de que o arquivo no padrão comma separated values (CSV) já havia sido processado e a informação disposta estava anonimizada[8]. Desta forma, para permitir sua ampla e irrestrita disseminação, as informações contidas nos registros impossibilitavam qualquer tipo de identificação dos pacientes. O arquivo obtido continha 564.163 registros, com informações sobre gênero, idade e município de residência dos pacientes, incluindo dados clínicos e sintomas apresentados, dados sobre o método de diagnóstico, entre outros. Considerando as informações existentes nos registros, o único critério de inclusão adotado foi a data do resultado do exame com confirmação de caso positivo da doença, de acordo com o período considerado na pesquisa. Considerando o município de residência informado nos registros existentes no conjunto de dados, adotou-se como critério de exclusão os registros com pacientes residentes em outros estados do país.

Na etapa de preparação (ou pré-processamento), os dados foram tabulados de modo a representar a incidência dos casos de COVID-19 em cada município, por data do resultado do exame de confirmação da doença. A preparação dos dados produziu uma série temporal com periodicidade diária, nos diferentes municípios de Santa Catarina. Partindo da premissa da existência de uma inter-relação epidemiológica, todos os municípios foram dispostos em uma tabela (ou conjunto de dados), mas em diferentes colunas, de modo a permitir uma modelagem de dados integrada que considerava essa conexão.

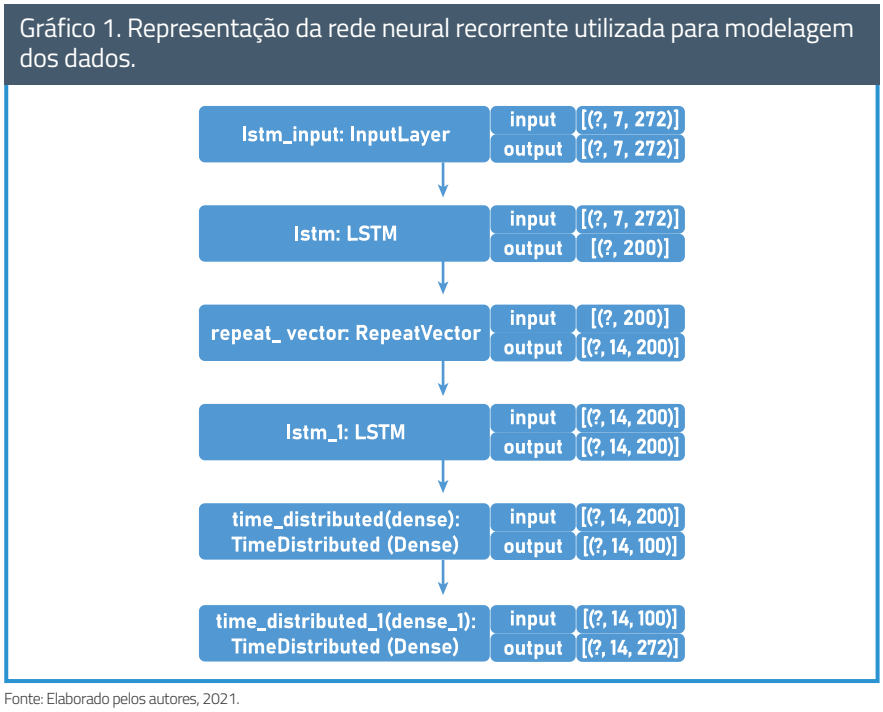
Com o conjunto de dados tabulados, uma análise de autocorrelação dos dados da série temporal foi também realizada através de funções de autocorrelação total e parcial. A análise heurística de todas essas informações possibilitou a determinação das fases seguintes à preparação dos dados. Para permitir a adequada avaliação do modelo de dados obtido, 30% da série temporal foi segregada como conjunto de dados de teste. Na modelagem de séries temporais, a relação entre as instâncias precisa ser conservada através da manutenção da sequência de disposição dos dados; assim, o conjunto de teste referia-se aos últimos dados disponíveis na tabulação anteriormente efetuada. Os dados de testes não foram utilizados nos processos de modelagem, permanecendo reservados apenas para validação dos modelos obtidos.

Em seguida, a modelagem dos dados e avaliação dos modelos foram executadas através do método walk-forward validation, utilizando o processo sliding window com o tamanho da janela (steps in) de entrada dos dados definida em 7 dias. Um método grid search foi aplicado para ajustar os hiperparâmetros do algoritmo de aprendizagem utilizado – epochs = 300, batch = 8 e nodes = 200 (com exceção da última camada intermediária que possuía metade da quantidade de neurônios artificiais). Os dados de saída do modelo eram as predições no horizonte de tempo de 14 dias, consideradas as variáveis dependentes utilizadas para análise da qualidade do modelo de dados através de uma métrica de avaliação.

Os processos de modelagem de dados foram realizados utilizando a biblioteca Keras[9], executada sobre uma API TensorFlow[10], em um ambiente de programação Python. Todas as etapas da pesquisa, incluindo o pré-processamento dos dados, a construção dos gráficos e as análises aplicadas nos modelos obtidos, também foram executadas no mesmo ambiente de programação. O código-fonte foi baseado em um programa de computador registrado no Instituto Nacional de Propriedade Industrial (INPI), sob o registro BR nº 512021000139-7[11].

A modelagem de dados foi realizada através de uma rede neural recorrente, utilizando múltiplas camadas Long Short-Term Memory (LSTM), que tem a capacidade de processar conjuntos de dados multivariados. Essa configuração consistia em um modelo encoder-decoder – composto por dois submodelos: um codificador e um decodificador –, proposto por Malhotra et al.[12] com algumas modificações sugeridas por Brownlee[13]. A rede neural recorrente aplicada no proces-

Gráfico 1. Representação da rede neural recorrente utilizada para modelagem dos dados.



Fonte: Elaborado pelos autores, 2021.

samento dos dados está representada no Gráfico 1. Para inicialização das matrizes de pesos foi aplicada a função desenvolvida por Glorot[14], com exceção das unidades de viés, nas quais foram aplicadas funções randômicas. Ambas as funções foram configuradas para obtenção de uma matriz de pesos com uma distribuição normal.

A métrica de avaliação root mean squared error (RMSE) – em português, raiz quadrada do erro médio quadrático – foi utilizada para avaliar os modelos obtidos comparando os dados de saída do processamento e os dados reais existentes, através da fórmula:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} (p_t - a_t)^2}$$

onde, pt é o valor predito e at é o valor atual, em determinado tempo t, para um conjunto de dados de tamanho n.

Os dados preditos pelo modelo foram comparados com os dados reais através de um gráfico, considerando o conjunto de amostras obtidas pelo método walk-forward validation, em cada dia do horizonte de predição. Os resultados obtidos foram confrontados com um modelo de persistência – que considera o valor da posição t+1 igual ao da posição t –, o qual é usualmente utilizado como linha de base para a performance de modelos de dados construídos por algoritmos de aprendizagem de máquina.

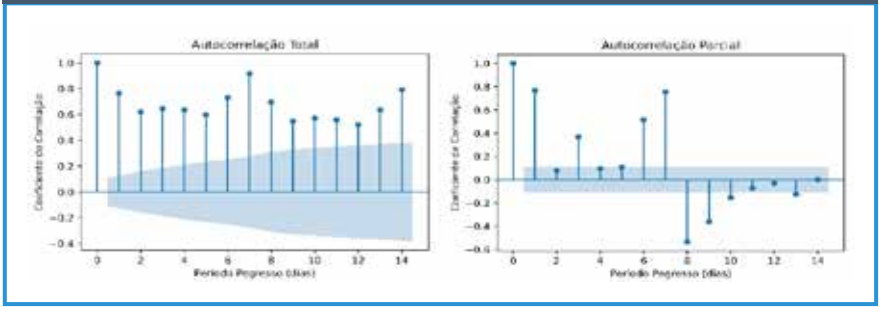
RESULTADOS

Com a tabulação do arquivo coletado no repositório do Estado de Santa Catarina, as funções de autocorrelação total e parcial confirmaram a existência de uma relação na sequência temporal, que permitiu inferir sobre a possibilidade de predição da progressão epidemiológica da doença. A aplicação da função de autocorrelação total demonstrou resultados estatisticamente significantes (intervalo de confiança de 95%) em todos os 14 dias pregressos, com o maior resultado no sétimo dia. A autocorrelação parcial apresentou resultados significativos em diferentes períodos de tempo, sendo um dos mais relevante ocorrido também em 7 dias pregressos; o qual, ao final, foi estabelecido como a melhor janela de entrada de dados para o modelo. O Gráfico 2 apresenta os resultados da aplicação das funções de autocorrelação total e parcial.

De acordo com a configuração da pesquisa, a rede neural recorrente LSTM empregada produzia uma sequência de predição de 14 dias para cada janela de entrada de dados de 7 dias. Desta forma, o RMSE foi calculado em cada uma das sequências de saída, entre os dados preditos pelo modelo e os dados reais do conjunto de dados de teste. Considerando todos os ciclos de modelagem executados pelo processo walk-forward validation, um RMSE médio de 20,74 foi obtido. O modelo de persistência, construído para possibilitar a comparação através do estabelecimento de uma linha de base, obteve um RMSE médio de 24,76. A Tabela 1 apresenta as informações estatísticas sumarizadas dos resultados de ambos os modelos, apontando ainda resultados menos dispersos no modelo de dados obtido (desvio padrão de 5,16) do que no modelo de persistência (desvio padrão de 6,85).

Também é possível analisar os resultados observando o Gráfico 3, que apresenta as saídas de dados do modelo de dados, em todas as janelas produzidas pelo processo walk-forward validation no conjunto de dados de teste. Cada um dos 14 gráficos representa todos os resultados consolidados de um determinado período do horizonte de tempo de predição (t+1, t+2..., t+14). Importante ressaltar que os gráficos individuais não expressam uma sequência temporal e sim, amostras de dados consolidados; contudo, esse formato possibilita a comparação dos valores reais do conjunto de dados de teste e os valores preditos pelo modelo de dados. Analisando os resultados dispostos no gráfico, é possível notar uma menor discrepância entre os valores reais e valores preditos no período de

Gráfico 2. Resultado das funções de autocorrelação total e parcial, no conjunto de dados tabulados.



Fonte: Elaborado pelos autores, 2021.

Tabela 1. Sumário estatístico dos resultados do modelo de dados, obtido pelo algoritmo de aprendizagem de máquina, e do modelo de persistência.

	MODELO DE DADOS	MODELO DE PERSISTÊNCIA
σ	20,74	24,76
σ	5,16	6,85
min	14,39	11,76
Q1	16,70	19,51
Q2	19,06	24,38
Q3	24,30	28,10
max	31,74	43,94

Fonte: Elaborado pelos autores, 2021.

tempo de um dia (t+1) do que o gráfico do último dia do horizonte de tempo (t+14).

DISCUSSÃO

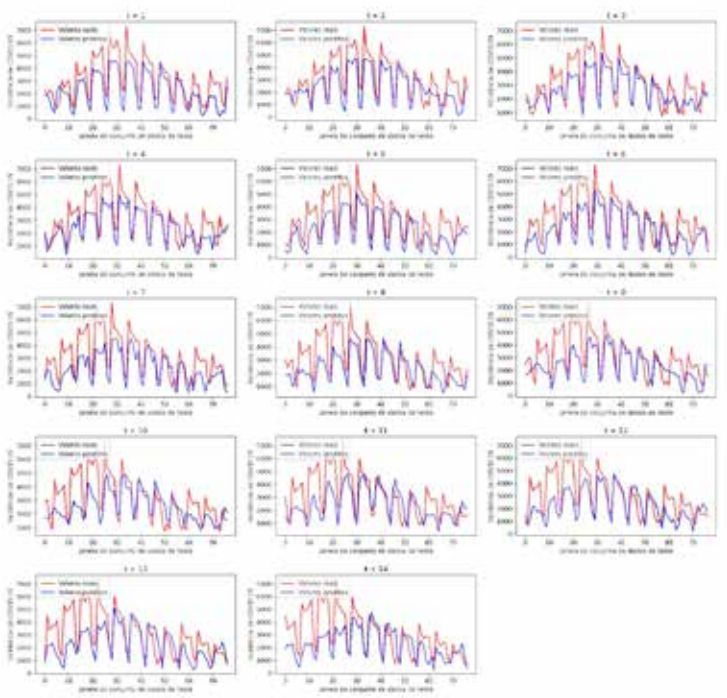
Dados tabulados em séries temporais são empregados em diferentes áreas do conhecimento, mas amplamente utilizados em estudos epidemiológicos. Séries temporais podem ser definidas como uma sequência de dados registrados em um intervalo regular de tempo, como, por exemplo, a incidência de uma doença[15]. Quando dispostos em gráficos cartesianos, a análise de séries temporais é tradicionalmente utilizada em estudos epidemiológicos, sendo também provedora de importantes informações e permitindo a construção de inferências úteis. Mas quando é necessário estabelecer inter-relações entre diferentes variáveis, outros métodos para a revelação desse conhecimento intrínseco devem ser aplicados. Nesse contexto, os algoritmos de aprendizagem de máquina, baseados nos fundamentos da Inteligência Artificial, sobressaem-se quando comparados ao método heurístico.

Considerando o limitado período de tempo analisado na pesquisa, restrito aos recentes meses da pandemia de COVID-19, foi necessário considerar alguns conceitos que proporcionassem a adequada modelagem dos dados utilizados na pesquisa. Conforme relatado por Mussumeci e Coelho[16], o uso de conjuntos de dados multivariados - no caso, caracterizado por uma série epidemiológica em diferentes municípios - permite explorar o conceito epidemiológico de similaridade. Consequentemente, a limitação relatada foi compensada pela amplitude dos municípios analisados e pela abordagem aplicada nos processos de pré-processamento dos dados. Sob outro aspecto, a premissa da existência de uma inter-relação epidemiológica foi determinante para a utilização das redes neurais recorrentes utilizada para a modelagem dos dados nessa pesquisa, pois permitem a entrada dos dados no formato adequado para modelar a similaridade epidemiológica, conforme pode ser observado na entrada (input) da primeira camada da rede neural recorrente apresentada no Gráfico 1 (com 272 municípios que contavam com casos de COVID-19 reportados).

Outrossim, redes neurais recorrentes possuem a capacidade de armazenar a representação de suas conexões anteriores, aplicando o modelo LSTM desenvolvido por Hochreiter e Schmidhuber[17]. Essa arquitetura permite modelar intervalos de tempo através de um fluxo de erro constante nas unidades de cálculo. Redes com camadas LSTM proporcionam o aprendizado de correlações de longo prazo em uma sequência, apresentando melhores resultados do que as redes neurais convencionais. Desta forma, principalmente na modelagem de sequências complexas multivariadas, as redes LSTM apresentam resultados mais precisos[12]. E apesar dos resultados apresentados considerarem todo o conjunto de dados, os modelos obtidos realizavam predições dos casos de COVID-19 em cada município, separadamente. Consequentemente, o modelo de dados pode ser aplicado para a predição individualizada, ainda que seja utilizado um período de tempo maior. Esse aspecto deve ser considerado importante, pois essa pesquisa pode produzir um instrumento útil no planejamento da intervenção sanitária durante a pandemia de COVID-19.

Em relação aos resultados obtidos, a tendência de aumento do RMSE é um comportamento esperado em razão do acúmulo erro ao longo do horizonte de predição, conforme o modelo gera predições mais distantes da janela de entrada de dados. A utilização de redes neurais recorrentes procurou minimizar esse problema, através do fluxo constante de cálculo do erro nas camadas LSTM. Sob outro aspecto, a inexistência de um ciclo sazonal na evolução do RMSE, ao longo do horizonte de predição, permite afirmar que foi possível modelar a sazonalidade semanal da série temporal analisada. A utilização de algoritmos de aprendizagem de máqui-

Gráfico 3. Valores reais (vermelho) e valores preditos (azul) da incidência diária de COVID-19, em diferentes períodos do horizonte de tempo de predição.



Fonte: Elaborado pelos autores, 2021.

na mormente desconsidera a necessidade de decomposição de séries temporais, que auxiliam a modelagem desses tipos de dados; porém, estudos futuros que apliquem o processo podem aprimorar os resultados dessa pesquisa.

Uma limitação inerente ao desenho da pesquisa proposta foi a ausência de segregação dos registros, considerando informações como, por exemplo, idade, gênero, e presença de comorbidades; ou, sob outros aspectos, a ocorrência de internação, incluindo em unidade de terapia intensiva, tipo de exame de diagnóstico realizado, entre outras informações contidas nos registros obtidos. A segregação pode conduzir a uma melhor análise dos vieses existentes nos dados analisados e ser abordada em futuros estudos epidemiológicos que detenham o objetivo de analisar essas características.

CONCLUSÃO

O resultado alcançado pelo modelo de dados, obtido através da aplicação de um algoritmo de aprendizagem de máquina, apresentou um erro médio menor erro do que a linha de base estabelecida através de um modelo de persistência. Consequentemente, conclui-se que a modelagem de dados executada permitiu a predição da incidência diária de COVID-19 em municípios de Santa Catarina, em um horizonte de tempo de predição de 14 dias. Considerando todos os aspectos relatados, a avaliação da predição de casos de COVID-19 possibilitou inferir que o emprego de ferramentas de Inteligência Artificial proporciona a obtenção de um importante instrumento para o enfrentamento da pandemia de COVID-19. Sua aplicação pode aperfeiçoar a capacidade de gestores públicos e administradores de instituições de saúde no gerenciamento de recursos, que precisam ser adequadamente alocados para uma resposta sanitária adequada ao avanço da doença. ■

REFERÊNCIAS

1. Freire-Silva J, dos Santos FH, Candeias ALB, Pinho MAB, Oliveira BRB. A utilização do planejamento territorial no combate da COVID-19: considerações sobre a situação dos leitos nos municípios de Pernambuco, Brasil. Vigil sanit debate. 2020;8(2):16-27. doi:10.22239/2317-269x.01546

2. Paumgartten FJR, Delgado IF, Rocha PL, de Oliveira ACAX. Drug repurposing clinical trials in the search for life-saving COVID-19 therapies. Vigil sanit debate. 2020;8(2):39-53. doi:10.22239/2317-269x.01596

3. Ministério da Saúde. O que é coronavírus? (Covid-19). Brasília, DF: Ministério da Saúde; 2020[acesso 16 ago 2020]. Disponível em: https://coronavirus.saude.gov.br

4. Lopes-Júnior LC. A Saúde Coletiva no epicentro da pandemia de COVID-19 no Sistema Único de Saúde. Saúde Colet. (Barueri). 2020;10(56):3080-9. doi:10.36489/saudecoletiva.2020v10i56p3080-3089

5. Bistrian DA, Dimitriu G, Navon IM. Processing epidemiological data using dynamic mode decomposition method. In: AIP Conference Proceedings; 2019 Jun 20-25; Albena, Bulgaria; Maryland: AIP Publishing LLC; 2019. doi:10.1063/1.5130825

6. Bontempi G, Taieb SB, Le Borgne Y. Machine learning strategies for time series forecasting. In: 2nd European Summer School Business Intelligence, eBISS; 2012 Oct 15-21; Brussels, Belgium; New York: Springer; 2012. doi:10.1007/978-3-642-36318-4_3

7. Wu Y, Yang Y, Nishiura H, Saitoh M. Deep learning for epidemiological predictions. In: 41st International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2018. 2018 Jul 8-12; Ann Arbor, United States. New York: Association for Computing Machinery. doi:10.1145/3209978.3210077

8. Controladoria-Geral do Estado de Santa Catarina. Portal de Dados Abertos do Estado de Santa Catarina. COVID-19 - Casos Confirmados. 2020[acesso em 14 ago]. Disponível em: http://dados.sc.gov.br/dataset/covid-19-dados-anonimizados-de-casos-confirmados

9. Chollet F, et al. Keras. 2015[acesso 20 mar 2020]. Disponível em: https://keras.io

10. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. TensorFlow: Large-scale machine learning on heterogeneous systems. [acesso 14 ago 2020]. Disponível em: https://www.tensorflow.org

11. Vianna, LS. Sistema de predição de casos de COVID-19 nos municípios de Santa Catarina. BR n° 512021000139-7. Concessão: 25 de fev. 2021.

12. Malhotra P, Vig L, Shroff G, Agarwal P. Long short term memory networks for anomaly detection in time series. In: 23rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2015; 2015 Apr 22-24; Bruges, Belgium.

13. Brownlee, J. Deep learning for time series forecasting: Predict the future with MLPs, CNNs and LSTMs in Python. Vermont: Machine Learning Mastery; 2018.

14. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics, JMLR Workshop and Conference Proceedings; 2010 May 13-15; Sardinia, Italy.

15. Bhaskaran K, Gasparrini A, Hajat S, Smeeth L, Armstrong B. Time series regression studies in environmental epidemiology. Int. J. Epidemiol. 2013;42(4):1187-95. doi:10.1093/ije/dyt092

16. Mussumeci E, Coelho FC. Machine-learning forecasting for Dengue epidemics-Comparing LSTM, Random Forest and Lasso regression. medRxiv. 2020[acesso em 20 mar 2020]. Disponível em: https://www.medrxiv.org/content/10.1101/2020.01.23.20018556v1. doi:10.1101/2020.01.23.20018556

17. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8): 1735-80. doi:10.1162/neco.1997.9.8.1735