



# Article COVID-19 Patterns in Araraquara, Brazil: A Multimodal Analysis

Dunfrey Pires Aragão <sup>1,2</sup>, Andouglas Gonçalves da Silva Junior <sup>3</sup>, Adriano Mondini <sup>4</sup>, Cosimo Distante <sup>2</sup> and Luiz Marcos Garcia Gonçalves <sup>1,\*</sup>

- <sup>1</sup> Pós-Graduação em Engenharia Elétrica e de Computação, Universidade Federal do Rio Grande do Norte, Av. Salgado Filho, 3000, Lagoa Nova, Natal 59078-970, Brazil
- <sup>2</sup> Institute of Applied Sciences and Intelligent Systems-CNR, Via Monteroni sn, 73100 Lecce, Italy
- <sup>3</sup> Instituto Federal do Rio Grande do Norte, Rua Dr. Mauro Duarte, S/N, José Clóvis, Parelhas 59360-000, Brazil
- <sup>4</sup> Faculdade de Ciências Farmacêuticas, Universidade Estadual Paulista "Júlio de Mesquita Filho", Rodovia Araraquara-Jaú, Km 1, Campus Ville, Araraquara 14800-903, Brazil
- \* Correspondence: lmarcos@dca.ufrn.br

Abstract: The epidemiology of COVID-19 presented major shifts during the pandemic period. Factors such as the most common symptoms and severity of infection, the circulation of different variants, the preparedness of health services, and control efforts based on pharmaceutical and nonpharmaceutical interventions played important roles in the disease incidence. The constant evolution and changes require the continuous mapping and assessing of epidemiological features based on time-series forecasting. Nonetheless, it is necessary to identify the events, patterns, and actions that were potential factors that affected daily COVID-19 cases. In this work, we analyzed several databases, including information on social mobility, epidemiological reports, and mass population testing, to identify patterns of reported cases and events that may indicate changes in COVID-19 behavior in the city of Araraquara, Brazil. In our analysis, we used a mathematical approach with the fast Fourier transform (FFT) to map possible events and machine learning model approaches such as Seasonal Auto-regressive Integrated Moving Average (ARIMA) and neural networks (NNs) for data interpretation and temporal prospecting. Our results showed a root-mean-square error (RMSE) of about 5 (more precisely, a 4.55 error over 71 cases for 20 March 2021 and a 5.57 error over 106 cases for 3 June 2021). These results demonstrated that FFT is a useful tool for supporting the development of the best prevention and control measures for COVID-19.

Keywords: COVID-19 dynamics; social distance; lockdown; time-series forecast

# 1. Introduction

*Sarbecovirus* SARS-CoV-2, the causative agent of the coronavirus disease known as COVID-19, was responsible for one of the most important health emergencies in the world. Infection with SARS-CoV-2 can be asymptomatic, present mild disease, or lead to upper respiratory symptoms and extrapulmonary syndromes that can cause death [1].

The virus spread occurs mainly through airborne transmission and contact with droplets and fomites [2,3]. In the absence of vaccines or other pharmaceutical interventions, the first control measures to avoid a growth in incidences were the use of masks and a reduction in community mobility. These non-pharmaceutical interventions (NPIs) were the initial strategies to reduce the basic reproduction rate of the disease (also known as  $R_0$ ) in different areas of the globe.

The relative success of these NPIs was demonstrated by their efficacy in controlling COVID-19 [4–6]. The deployment of NPI measures helped to delay the massive growth in cases in nations where such policies were adopted. The use of face masks, along with social-distancing measures, were the first control strategies to restrain SARS-CoV-2 transmission [7,8]. In fact, these fast, standardized, and cost-effective countermeasures were our only line of defense against the spread of COVID-19 incidences until the use of vaccines was approved [9].



Citation: Aragão, D.P.; Junior, A.G.d.S.; Mondini, A.; Distante, C.; Gonçalves, L.M.G. COVID-19 Patterns in Araraquara, Brazil: A Multimodal Analysis. *Int. J. Environ. Res. Public Health* **2023**, *20*, 4740. https://doi.org/10.3390/ ijerph20064740

Academic Editor: Fernando Augusto Lima Marson

Received: 26 January 2023 Revised: 3 March 2023 Accepted: 4 March 2023 Published: 8 March 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). Different studies have shown the positive effect of social distancing and the restriction of human mobility in controlling the spread of SARS-CoV-2 [10–12]. However, NPIs such as social distancing presented a major impact on poverty and inequality [13]. The epidemiological value of such interventions was less evident in lower-income countries, as they imposed a heavy burden on vulnerable groups [14,15]. Another important aspect of NPIs and COVID-19 itself is the toll on mental health that may have affected the perception of risk as well as the importance of individual and collective control measures [16–19]. Therefore, the investigation of the potential outcomes related to the full and partial implementation of NPIs is of paramount relevance for COVID-19 epidemiology.

In Brazil, the main reason for the adoption of NPIs was the increasing number of cases in 2020–2021, especially in settings where no measures to avoid crowding were implemented [20]. At that time, multiple cities suffered from an insufficiency of ICU beds and a shortage of healthcare workers and supplies, which elevated the number of cases and deaths [21,22]. In such a scenario, cities with a smaller gross domestic product and modest health equipment structures struggled to manage the increase in healthcare demand due to SARS-CoV-2 circulation. The failure of the country to implement effective control measures, the denial of the pandemic's severity, and the use of inefficient early COVID treatments had a direct impact on the number of cases [23].

During 2020 and 2021, Brazil was under the government of the far-right president Jair Bolsonaro, whose administration advertised non-evidence-based pharmacological interventions and disregarded dependable public health measures such as the use of face masks and social distancing [23]. These actions also impaired the implementation of public health services and policies at the federal level and in other spheres of administration. During this period, the government made concerted efforts to demobilize analytical skills as a crucial component of the National Healthcare System [24]. Such strategies exposed, in particular, the somewhat unexpected size of the political bias regarding the spread of SARS-CoV-2 and the mortality rate of COVID-19 in Brazil, surpassing other important epidemiological elements such as the level of poverty and the mutation of the virus. People were constantly persuaded not to adopt restrictive measures [25,26]. Furthermore, the fomentation of anti-COVID-19 measures by the federal government amplified the challenges of monitoring the spread of the virus, to the extent that the daily count of cases and deaths was not properly updated. Such coordinated actions influenced decision-making processes in several other spheres of administration. Consequently, health managers were unaware of the true scope of the epidemic, which may have hindered the deployment of control measures at key points in the spread of SARS-CoV-2 [27].

States and cities run by political opponents of president Jair Bolsonaro deployed NPI measures independently of federal sanitary orientations [28]. An example was the city of Araraquara, São Paulo, Brazil. This city has an estimated population of approximately 238,000 people and a demographic density of 235 inhabitants/km<sup>2</sup>, according to IBGE [29]. Local authorities implemented different degrees of restrictive NPI measures, the red, yellow, and green flags, in response to the severity of the epidemiological scenario. Araraquara deployed strict social-distancing and lockdown measures to avoid the collapse of the health system as the cases of COVID-19 started to increase. The so-called red-flag measures limited certain economic activities and sectors but did not promote the complete shutdown of services, which happened when lockdown measures were implemented. Araraquara also issued two lockdown orders in 2021, one in February and another in June. As a result, there was a 43.02% decrease in cases after the first ten days of these interventions. Additionally, the time frame comprising March and June of 2021 presented more cases than reported in the previous 10 months [30]. As one of the largest cities in the region, Araraquara stood out nationally for implementing complete lockdown strategies and carrying out mass testing for COVID-19, which culminated in a decrease in death rates associated with the implementation of vaccination strategies in 2021 [31].

The diversity of the epidemiological patterns of COVID-19 during the pandemic was a result of the circulation of different SARS-CoV-2 variants [32]. Thus, it is important to

evaluate the evolution of common symptoms; the influence of environmental indicators such as temperature, humidity, and air quality on the spread of the virus; the effect of events such as vaccination and crowding; sociopolitical actions in neighborhoods or cities; and the type of measures implemented to control the spread of the virus, which are among the factors whose patterns change throughout pandemics, seasonally or non-seasonally.

In order to identify and understand some of these disease patterns, researchers have demonstrated the association of COVID-19 cases with crowding and NPI implementation [8,9,33–35]. In this sense, COVID-19's spread is regarded as a multifaceted sociopolitical problem, for which exploratory analysis and data comparison based on urban mobility and sociodemographic indicators can provide useful epidemiological information [36–40].

Hence, in order to better understand the pandemic scenario in Araraquara and what the forecasting of COVID-19, we applied mathematical and epidemiological models in neural networks to forecast the number of cases and observe the possible effects of restrictive measures in Araraquara, Brazil. Our assessment was based on previous studies that also used urban mobility and temporal and quantitative epidemiological data. Furthermore, by consulting the datasets thoroughly, we outlined the COVID-19 epidemiological scenario in Araraquara in 2020 and 2021, as well as the community mobility and its potential contribution to disease transmission. In addition, we used an initial Fourier transformation application to identify the disease's seasonal spread despite the implementation of control measures, which demonstrated that there were seasonal events to which specific NPIs could have been applied in certain key periods.

# 2. Materials and Methods

In this section of the manuscript, we provide a comprehensive description of the datasets and methods used for the analysis. We also highlight important hallmark dates that may have influenced the epidemiological patterns of COVID-19 in Araraquara.

#### 2.1. Dataset

In order to better comprehend the phenomena, we compared information from three different data sources:

- Community mobility reports (CMRs) from Google to analyze social dynamics;
- Municipal daily reports (MDRs) from bulletins provided by local authorities;
- Health department reports (HDRs) from local health authorities.

The CMRs [41] comprise data collected from tracked mobile phones/smartphones. The entries are associated with the activities and mobility of users in different sites and environments such as work, residential neighborhoods, and public areas. The CMRs are divided into six categories: retail and recreation, parks, groceries and pharmacies, workplaces, public transit, and residential. We collected mobility entries from February 2020 to June 2021. The CMRs include different levels of geographical organization, such as country, state, and city. The analysis of this dataset provided inputs of community mobility dynamics and local incidences during the red-flag and lockdown periods in Araraquara.

The COVID-19 epidemiological data present in the MDR dataset were obtained from daily bulletins provided by the local government [42]. This dataset is organized as a time series of cases, with the first registry dated 3 March 2020, and all subsequent data added daily. The dataset includes information such as the date of the report, Brazilian province, city, number of cases, number of deaths, estimated population, city code, percentage of cases per 100,000 inhabitants, and death rate.

The Health Department of Araraquara maintained a weekly updated database (HDR dataset) that was partially used in our analysis. The dataset comprised reports of 77,809 RT-PCR and rapid tests, including antibody or antigen test, and additional variables collected from 1 April 2020 to 31 May 2021, and was composed of 26 columns, organized as follows: identification column; three columns containing report date, date of onset of symptoms, and type and date of diagnostic test; one column related to the test result; one column

identifying whether the user was a health professional; and other columns related to the patient's chart, such as possible diseases and symptoms presented (see Table 1).

Table 1. Health department reports (HDRs) data columns.

Report number	Sore throat	Obesity
Report date	Dyspnea	Coryza
Date of onset of symptoms	Fever	Asymptomatic
Test date (PCR/rapid)	Cough	Race/color
Result (PCR/rapid)	Others	Pregnant
Olfactory disorders	Headache	Diabetes
Chronic heart disease	Taste disorders	Puerperium (up to 45 days after labor)
A healthcare professional?	Carrier of chromosomal diseases or immunologically fragile state	Advanced-stage chronic kidney disease (grades 3, 4, or 5)
Uncompensated chronic respiratory diseases		

Finally, we considered important dates for our analyses, as organized in Table 2. The first dates were the national elections, which occurred on 15 and 29 November 2020, resulting in crowding. Additionally, Araraquara's government implemented NPIs (red-flag alerts) during 2020 on 23 March, 1 May, 4 September, and 22 December. They also implemented NPIs on 2 April 2021. Further, they declared a lockdown on 21 February 2021, spanning to 2 March, and from 20 June to 27 June. Some of these dates explain the behavior of the temporal trends in Araraquara, as will be shown below.

#### Table 2. Calendar of main events in Araraquara.

ockaown
ebruary 2021 June 2021

## 2.2. t-Distributed Stochastic Neighbor Embedding (t-SNE)

The community mobility reports (CMRs) dataset provided a complex arrangement of mobility that evolved through many community sectors and changed over time when local events occurred. To generate independent variables capable of generalizing a temporal event or circumstance and simplify the data interpretation and visualization by reducing dimensionality through projecting the data to a lower-dimensional domain and preserving local structures, we used the t-Distributed Stochastic Neighbor Embedding (t-SNE) method proposed by Van der Maaten and Hinton [43], based on Stochastic Neighbor Embedding (SNE) [44]. However, there was a trade-off between precision (achieved through high dimensionality) and interpretability (given low-throw dimensions).

We identified works that used different methods for this task, such as those that used the PCA method, including the approach presented by Aragão et al. [34]. We adopted t-SNE, however, because it could take a high-dimensional dataset and reduce it to a lowdimensional graph while retaining the original information and meaning in the data. PCA (Principal Component Analysis) is a method for reducing the number of variables (generating principal components—PCs) in order to generalize the projections without losing substantial information [45]. To set up the sorting PC, the algorithm measures the average value of all dimensions of the data to avoid an unequal contribution from a given dimension into the process and applies a standardizing method to generate an input in a standard scale. Afterwards, the covariance of the variables is computed, and the covariance matrix *A* is established, followed by the use of a linear transformation to find the characteristic vectors (also known as *eigenvectors*). The chosen PC variables are the *eigenvectors* with the highest *eigenvalues*. The results are presented using an orthogonal axis designed to explain the maximum amount of variance in the set, which may result in information loss.

The t-SNE method worked by first calculating the Euclidean distances between points on the dataset, which were then translated into conditional probabilities that reflect the similarity between each pair of sites using:

$$p_{i|j} = \frac{exp(\frac{-|x_i - x_j|^2}{2\sigma_i^2})}{\sum_{k \neq i} exp(\frac{-|x_i - x_k|^2}{2\sigma_i^2})}$$
(1)

in which the division represents each possible cluster with different densities.

1

The next step was to create a random dataset in low-dimensional space and compute a joint probability distribution based on k-features (our target, in general, was 2 dimensions as the outcome). We established a joint probability distribution for this set of points, but at this stage utilized the t-distribution rather than the Gaussian distribution, as was the case in the initial dataset, adopting:

$$\eta_{i|j} = \frac{(1+|y_i-y_j|^2)^{-1}}{\sum_{k \neq l} (1+|y_k-y_l|^2)^{-1}}$$
(2)

Finally, we used the Kullback–Leiber (KL) divergence measure to get as near to the original dataset's joint probability distribution of data points in the low-dimensional space as possible. The value of the KL divergence decreases as the distributions become more similar, eventually reaching 0 when the distributions are identical. The KL divergence is defined as:

$$D_{KL}(P|Q) = \sum_{x \in X} P(x) log\left(\frac{P(x)}{Q(x)}\right)$$
(3)

In the approach presented in this work, we applied this method to the CMR dataset. However, we removed the component that represented *residential* mobility, keeping only those components relating to the different interactions across the city.

## 2.3. Fast Fourier Transform—FFT

Certain events occur in a seasonal manner, and COVID-19 is influenced by seasonal events. Two examples are the temperature and humidity indexes, which are seasonal and have a relation to the dynamics of COVID-19 [33,46,47]. The Fourier transform (FT) may help us comprehend the construction of the COVID-19 curve using an alternative method.

First, a periodic signal can be represented by a Fourier series, which is an infinite sum of sinusoidal functions (cosine and sine), each with a frequency that is an integer multiple of f = 1/T:

$$f(t) = a_0 + \sum_{m=1}^{\infty} a_m \cos^{\frac{2\pi m t}{T}} + \sum_{n=1}^{\infty} b_n \sin^{\frac{2\pi n t}{T}}$$
(4)

where  $a_0$ ,  $a_m$ , and  $b_n$  represent the Fourier series coefficients, which determine the relative weights for each of the sinusoids.

In order to depict a function as a continuous or integral superposition of complex exponential functions, we can then use the Fourier transform. A series can thus be represented in the frequency domain. To perform this method, we used decomposition, which converts (or extracts) the signal as a sum of other signals or as the sum of a series of *n*-sine waves represented as continuous signs.

In the case of non-continuous problems with information about a set of data rather than a continuous function, the discrete Fourier transform (DFT) could assist in transforming this dataset into an equal-sized set with information about the frequencies of the function that satisfies the dataset, or it could provide the discretized form of the Fourier transform. An arbitrary function that is periodic on a certain domain must be projected into orthogonal function directions by expanding it as a sum of sines and cosines as well as periods on that domain. For the 1-dimensional DFT (y[k]) of length N, we used the following definition:

$$y[k] = \sum_{n=0}^{N-1} x[n]e^{-2\pi nk/N}, k = 0, \dots, N-1$$
(5)

where the daily number of SARS-CoV-2 cases reported on the  $n_{th}$  day of the time series is represented by x, and y[k] denotes the magnitude of the  $k_{th}$  frequency on the n-day of the time series. By converting the time series of the number of new daily SARS-CoV-2 cases to the frequency domain, we used the DFT to estimate the period's length in the COVID-19 data spectrum.

To compute the discrete Fourier transform of a signal, a computer needs to complete N multiplications × N additions, resulting in an algorithm with a complexity of  $O(N^2)$  operations. The fast Fourier transform (FFT), as the name indicates, is a method that finds the discrete Fourier transform of an input much quicker than computing it directly. The FFT algorithm decreases the number of calculations required for a problem of size N from  $O(N^2)$  to O(NlogN), using the so-called *divide-and-conquer* approach. Rather than working with big signals, the algorithm breaks them into smaller ones and applies the DFT to these, before summing all of the smaller DFT results to obtain the final outcome, asymptotically providing a significant advantage. Hence, the FFT is an algorithm that allows one to quickly transform a series of discrete signals into a sample containing the frequencies of these signals, as long as certain properties are met.

Given this advantage, the intent was to consider the time series of COVID-19 cases extracted from the MDR dataset. The FFT allowed us to extract and map event frequencies for the city of Araraquara, displaying seasonality throughout the year. The frequencies were determined by the function and reported in a graphical form using four categories: (1) those representing events with seasonality ranging from 9 to 12 months (40–52 weeks); (2) those representing events with seasonality ranging from 6 to 9 months (26–40 weeks); (3) those representing events with seasonality ranging from 3 to 6 months (12–26 weeks); and (4) those representing events with seasonality ranging from 1 week to 3 months (1–12 weeks).

If a frequency peak appeared in a specific sector, such as between 3 and 6 months, it meant that the city had a high rate of COVID-19 cases, which repeated itself in a 3–6 month period. This method may enable one to map the events that correspond to each seasonality frequency and identify the best times to implement a measure, in addition to understanding its magnitude in the context of case control reporting.

#### 2.4. Seasonal Auto-Regressive Integrated Moving Average—SARIMA

The ARIMA and SARIMA methods are used in several applications, especially in univariate tasks. It is possible to refer to many works that used these methods to perform regression during the COVID-19 pandemic [34,48–52]. The methods consist of combining auto-regressive (AR) and moving-average models (MA) and then defining the ARIMA(p,d,q) model, allowing one to accommodate different types of non-stationary time series (*d* is the order of differentiation applied for the time series).

AR(p) is a *p*-order auto-regressive model that is based on a linear combination of past observations  $X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + ... + \alpha_p X_{t-p} + \varepsilon_t$ , where  $\alpha$  is constant and  $\varepsilon_t$  is a stochastic disturbance from the white noise process. When adjusting the parameters of an auto-regressive model, the parameter  $\alpha_k$  must be estimated, resulting in the equation:

$$Y_t = \mu + X_t = \mu + \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} + \varepsilon_t$$
(6)

On the other hand, MA models are an extension of the white noise process, consisting of a linear combination of the disturbance  $\varepsilon_t$  added to recent disturbances  $\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, \varepsilon_{t-q}$ . Thus, a *q*-order MA model can be defined as follows:

$$X_t = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \ldots + \beta_q \varepsilon_{t-q}$$
(7)

Seasonal Auto-regressive Integrated Moving Average (SARIMA), also known as Seasonal ARIMA, was created to support univariate time series based on ARIMA models. It contains a second set of parameters  $(P, Q, D)_s$  referring to seasonal auto-regression in the seasonal period *s*.

The advantage of using the SARIMA model is that there are fewer hyper-parameters, making the configuration file readily manageable if the model is put into production. However, when p and q rise in value, there are more coefficients to fit, increasing the time complexity exponentially. If p and q are large, the procedure becomes difficult to fit, depending on the complexity of the dataset.

## 2.5. Transformers

Certain methods are useful for situations that cannot be addressed linearly. In such circumstances, neural networks allow one to detect the degree of links among variables, such as those in our study, because non-linear variables cannot be described. Hence, neural networks show promise for univariate, bivariate, and multivariate regression issues.

One possible approach is to use the Transformer Multi-Head Attention model (or simply the Transformer model), which excels at managing sequential data, allowing one to tackle sequential and temporal topics without being limited in prediction. The core of the *Transformer* model's architecture is based on attention mechanisms [53]. Transformers have demonstrated outstanding performance in natural language processing and (especially) in computer vision, which the scientific community has exploited in the time-series domain [54]. This method has the advantage of capturing long-term data dependencies and relationships and producing better data visualizations.

The Transformer model's outermost structure consists of an encoder and a decoder, and the core is built on self-attention method processes. When adding positional encoding to the input and output sequence embedding, self-attention allows the model to search at other points in the input sequence for hints that might assist in better encoding a target sequence. This approach uses positional encoding appended to input embedding to describe the input series representation, with no recurrence or convolution operations. The inputs to both the encoder and decoder components use the same embedding and positional encoding logic, taking both sequences of input and output tokens and converting them into vectors with no order (a query vector, key vector, and value vector for each input). Afterwards, a component of the embedding vectors called the positional encoding converts the sequence to sine and cosine frequencies. In this case, nearby elements have a similar positional encoding.

Each encoder block has a multi-head self-attention module and a position-wise feedforward network (FFN), while each decoder block inserts cross-attention models between the multi-head self-attention module and the position-wise feed-forward network. The multi-head attention mechanism uses the values, keys, and queries in the so-called Scaled Dot-Product Attention process to calculate the attention weights based on:

$$Attention(Q, V, K) = softmax(\frac{QK^{T}}{\sqrt{d_{K}}})V$$
(8)

where Q, K, and V are *queries*, *keys*, and *values*, respectively, and  $d_K$  is the dimension key, which are used to regularize the result with a softmax function. The softmax function layer then turns these scores into probabilities (all positive, adding up to 1.0). The cell with the highest probability is chosen, and the input associated with it is identified as the output of the time step.

The network adopted in this work had a head size of 200 blocks, with 4 heads containing transformer blocks connected to an MLP with 128 units and a dropout of 30%. We used the univariate model based on COVID-19 daily cases as the input data to predict two different outcomes: the next day following a true sequence, and the next day following a sequence that was constructed based on each sequential prediction. The learning algorithm was asked to output the function  $f : \mathbb{R} \to \mathbb{R}$  in order to complete this task.

#### 3. Experiments and Results

We started by extracting information based on graphs that were constructed for data from the MDR, HDR, and CMR datasets during 2020 and 2021.

## 3.1. COVID-19 Patterns in Araraquara

From the HDR dataset, we identified 77,809 records, and the overall sum of positive tests (including RT-PCR and fast tests) was 14,729; tests that presented negative to COVID-19 comprised 47,432 records. Additionally, 52 sample records were designated as inconclusive or uncertain. From the 14,729 COVID-19-positive tests, 1838 (approximately 12.5%) were reported as individuals with symptoms, whereas 12,891 (approximately 87.5%) were recognized as COVID-19-positive individuals who were asymptomatic and did not report symptoms.

The most frequently reported symptoms among the positive-RT-PCR patients were cough (reported by 41.16%), fever (30.06%), sore throat (29.30%), and headaches (27.69%), as shown in Figure 1. Additionally, we observed that when users were identified as positive, the symptoms related to olfactory, gustatory, and dyspnea disorders were the least reported.



**Figure 1.** Symptoms reported by suspected (blue) and confirmed (orange) COVID-19 patients from Araraquara, São Paulo, Brazil, from March 2020 to July 2021.

Among those individuals identified as positive cases, 2456 had some condition or comorbidity that characterized them as a special-attention case or was considered a risk factor (Figure 2). This group was distributed under the following classifications according to the individual's condition: obesity; postpartum or pregnant women; patients with chronic respiratory, cardiac, renal, or chromosomal diseases or frailty; and patients with diabetes or immunosuppression.

In relation to the importing and exporting of cases, the movement of the disease between cities within the state of São Paulo and its neighboring states had a significant influence. For example, 7888 out of the 77,809 case records identified as Araraquara residents were imported from cities in other states, such as Curitiba, located in the neighboring Paraná state. The distance between Araraquara and Curitiba is approximately 664.3 km.



**Figure 2.** Types of comorbidities and health conditions presented by residents of Araraquara, São Paulo, Brazil, comprising cases reported from March 2020 to July 2021.

The state of São Paulo borders four other states: Rio de Janeiro, Minas Gerais, Paraná, and Mato Grosso do Sul. Figure 3 depicts the cases imported and exported to/from other states, including Minas Gerais and Rio de Janeiro. These three states are the most populated states in Brazil according to the IBGE [29], with São Paulo representing 21.9% of the Brazilian population, Minas Gerais 10.1%, and Rio de Janeiro 8.2%. Considering the states with more than 80 instances imported or exported to another state, the emigration from Paraná to São Paulo represented 7664 records, and that from Minas Gerais and Rio Grande do Sul to São Paulo accounted for 104 and 94, respectively. Emigration from São Paulo to Minas Gerais accounted for 89 of the instances. The remaining cases for São Paulo originated in the state itself (69,442 cases).



Figure 3. Autochthonous and imported reported cases, state of São Paulo, Brazil, 2020–2021.

#### 3.2. Community Mobility and Frequency Events

It was possible to extract event-based information for 2020 and 2021, in particular, in relation to the variation in COVID-19 daily cases. To this end, we used the MDR dataset that is graphically represented in Figure 4. In this graph, for ease of understanding, we applied a moving average of 7 days and included all official public holidays in Brazil as red lines. Additionally, the events presented in Table 2 (Section 2) were included, i.e., national elections, NPIs, and lockdown measure implementations.

Lockdown periods are represented as blue stripes, indicating the exact length in relation to the timeline and the determined period span. Green stripes correspond to NPI measures, representing the first day of the measure's implementation plus the seven days following. We adopted this approach as it could make it easier to observe how these measures graphically demonstrated their outcomes in the period of 7 days. As shown in Figure 4, these types of measures, such as NPIs and lockdowns, were implemented when



the curve's trend was at its local peak (when considering a short period, this could be determined as the peak of cases in the selected period).

**Figure 4.** Cases reported (blue line) by health authorities from Araraquara, São Paulo, Brazil, from March 2020 to July 2021.

There was a local peak between August and September 2020. During this period, the government of Araraquara authorized bars and restaurants to operate, which may have contributed to the increasing number of cases. Given the increasing number of cases during other periods, we assumed that the purpose of implementing these measures was to cause a recession or spread stagnation. Furthermore, we observed that at least half of the holiday dates appeared to precede an increase in daily cases, contributing to the upward trend of the curve. This was visible during periods such as the holidays preceding the March 2021 lockdown (03-21 in the graph). Some holidays in 2021 followed the same pattern, such as those in April and June (04-21 and 06-21 in the graph). It is important to note that the lockdown periods were implemented immediately after the highest rates of COVID-19 daily cases in Araraquara.

We used data from the CMR database and ran them through the t-SNE method to help us understand how these high numbers of daily COVID-19 cases came to be. However, first, we verified if there was a correlation between the CMR features (Figure 5). With the exception of transit stations, residential mobility had an inverse connection with all other components. The other elements were positively connected to one another. Furthermore, the strongest negative correlation was between the residential and workplace features.



Figure 5. CMR feature correlation.

Furthermore, before employing t-SNE, we conducted a Shapiro–Wilk test [55] to see if the distributions were close to normal distributions (Figure 6). All CMR characteristics had *p*-values equal to 0.000, leading us to reject the null hypothesis  $H_0$  (that the data would follow a pattern of development) and indicating that the data did not display Gaussian distribution behavior.



**Figure 6.** Shapiro–Wilk outcomes: (**a**) Residential. (**b**) Parks. (**c**) Transit Stations. (**d**) Groceries and pharmacies. (**e**) Workplaces. (**f**) Retail and Recreation.

The t-SNE method could help us reduce the number of community mobility components analyzed from five (retail and recreation, parks, groceries and pharmacies, workplaces, and public transit) to two, as shown in Figure 7. The algorithm looked for similarities between all the points based on their distance and divided them into two groups, which we will refer to as Component 1 and Component 2.

Comparing the components generated by the t-SNE method and the holidays, NPIs, and lockdown measure implementations, we could determine that:

- The first NPI was adopted one month before the first wave of cases in the city.
- The second and third NPIs were implemented at a time of a high and stable number of cases.
- The mobility behaviors presented in the first and second lockdowns were not similar. The first lockdown demonstrated a greater effective mobility restraint force compared to the second.

- Comparing with Figure 4, it was observed that the mobility restrictions could reduce the number of cases but were not so efficient as to decrease the number of daily cases to zero.
- The rises in the curves of the urban mobility components, in general, were preceded by a holiday.

We then applied the FFT in an attempt to specify if there were periods during which events were distributed that influenced the case curve. It is important to note that in this case, the FFT method was only used to identify the Fourier frequency spectrum, and the method was applied to the original temporal record of cases, i.e., the series was applied without using a moving average.



Figure 7. t-SNE Components.

Figure 8 depicts a frequency spectrum with endpoints at junctions that correspond to COVID-19's case peaks. It was necessary to convert this to a semi-logarithmic annual scale, in which each peak is represented by its magnitude (on the *y*-axis) and *k*-frequency (on the *x*-axis). Without the 7-day moving average treatment, the DFT was applied to the original curve.



Figure 8. COVID-19 FFT frequency events.

The graph shows a frequency peak for the 7-day cycle associated with the case notification process (with accumulated cases reported on Sunday and Monday), which was repeated 52 times a year (k = 52), where 52 is the number of weeks in a year. Furthermore, we discovered another point with a high magnitude on the extreme left side of the spectrum that was not categorized as a peak, which was a phenomenon that occurred every 6 to 9 moths of the year during the observation period. This implied that the repeat cycle was roughly k = 1.55. This phenomenon could be identified as the presence of so-called popular waves, defined as a sudden increase in the number of cases compared to the city's average. Furthermore, three other phenomena or events recurred over a period of 3 to 6 months (k = [2.33, 3.12, 3.9]). Except for the event at k = 52, we could identify this as a single moment that appeared to be repeated every 1 to 3 months, that is, with k = 5.46.

# 3.3. Forecasts

In the following, we present some predictions that could support and help in determining if it is possible to introduce the hypothesis that lockdowns were an important measure to contain or reduce the COVID-19 daily cases in Araraquara.

# 3.3.1. SARIMA

The first regression approach for the two lockdown measure periods used the SARIMA model. In the training step, it was observed that the model maintained the curve's trend and its patterns when compared to the ground truth throughout the predictions, as shown in Figure 9.



Figure 9. SARIMA model forecasts for the first lockdown from 21 February to 2 March in Araraquara.

Figure 9 presents the results from a trained SARIMA model. The base data for the training step ranged from the first reported case (25 February 2020) to the day of the first lockdown (21 February 2021). The model parameters—included in the definition  $SARIMA(p,d,q)(P,D,Q)_s$ —were determined using the Box and Jenkins approach [56,57]. The model applied for the first lockdown period was defined in the order (0, 1, 0), with the seasonality order (7, 1, 0, 7).

We also considered that an infected person's symptoms begin to appear approximately 4 or 5 days after infection, as presented by Johansson et al. [58]; hence, if the transmission took place on day d, we could observe its results on day d + 7 [1,59]. Given this, when choosing the period for prediction (the data used in the test), we considered that results could occur within a timespan of 6 days after the first day of lockdown implementation, representing the future outcome of the implemented measures in terms of COVID-19 daily cases.

The predictions made by the model in the test step demonstrated that it was expected that the curve's trend and the number of cases would increase. The real number of cases in 20 March 2021 was 71. For the same date, the model predicted a value of 166.17, representing a value difference of approximately 94 cases. Most importantly, the RMSE extracted based on the sweeping curve was RMSE = 4.55 cases.

Furthermore, the model was trained for the second lockdown period and defined in the order (0, 1, 0), with the seasonality order (8, 1, 0, 7). The predictions for the second lockdown are shown in Figure 10.

We observed different patterns for the curve's trend and different COVID-19 daily case expectations in this prediction set. In contrast to the first lockdown, the forecasts showed that the number of cases would remain stable, with no significant increases. This result

may have been based on the previous data pattern curve, which was presented above for the beginning of that timespan.

The real number of cases in 3 June 2021 was 106. The model predicted a value of 97.09 for the same date, representing a value difference of approximately 9 cases. However, the model showed a significant value divergence for previous dates in that period. The RMSE extracted based on the complete curve was RMSE = 5.57 cases.



Figure 10. SARIMA model forecasts for the second lockdown from 20 to 27 June in Araraquara.

# 3.3.2. Transformer Model

Next, we present the predictions made using the Transformer models. We decided that, in order to predict one day, a sequence of seven days had to be presented, which meant that in order to predict the eighth day, we had to first present data corresponding to seven consecutive days of COVID-19 cases. The model received a sequence of data in both the training and testing stages, forming a vector of information. Similarly, to predict day 9, we presented a 7-day sequence that corresponded to days 2–8. To forecast day 10, we presented data from days 3 to 9.

Figure 11 depicts the actual COVID-19 daily cases, those predicted by the model for the same timeline, and the forecast for cases immediately after the lockdown was implemented on 21 February 2021. This last predicted series was obtained with the assumption that the forecast day should be included in the following time series used with the forecast. In this case, suppose we use the real series from days 1 to 7 to forecast the next day (eight days):  $[1_d, 2_d \dots 6_d, 7_d] \rightarrow 8_p$ . Once the forecast for day 8 is complete, its value is attached to the next series as a data entry, which means that for day nine, we consider days 2 through 7 plus the previously predicted eight day:  $[2_d, 3_d \dots 7_r, 8_p] \rightarrow 9_p$ .



Figure 11. Transformer model forecasts for the first lockdown from 21 February to 2 March in Araraquara.

As observed, the pattern learned by the model demonstrated that the predicted data expected that the trend in the number of cases would continue to rise, although not as sharply as the current data presented. This could be explained by hidden information that was not presented, such as the fact that crowding had occurred in the previous period, which could be critical for the model to determine a trend more effectively and accurately. As with the SARIMA model's predictions, once real temporal information was not provided, the model predicted an increase in the number of cases due to COVID-19. At the test stage, the model had an RMSE of 13.30. The sum of COVID-19 cases during the test period was 2783, while the sum of values predicted by the model for the same time period was 2878.

Similarly, Figure 12 demonstrates the same pattern in predicting the days that occurred after the first lockdown. In this second timespan, the pattern diverged from the predictions made by the SARIMA model. This showed that the restriction measure's implementation had a significant impact, preventing the recognized standard expectation of an increase in cases due to the oversaturation of the health system. In this process, the model presented a slightly higher RMSE than in the previous stage, with a value of 14.39 in the test step. For the test data, the total number of COVID-19 cases was 1746, whereas the sum amount predicted by the model was 1745.



Figure 12. Transformer model forecasts for the second lockdown from 20 to 27 June in Araraquara.

#### 4. Discussion

In the present study, we were able to assess the epidemiological profile of COVID-19 in a medium-sized city in Brazil, applying mathematical models and neural networks to analyze disease reports, community mobility, and NPIs. The ability of the models to properly forecast disease patterns in Araraquara supported their use as potential algorithms for situation diagnosis in different epidemiological scenarios. Additionally, our data strongly suggested that lockdown measures were of paramount importance in decreasing COVID-19 incidences in the city after only a few days of implementation.

It is not possible to state that the only factor that influenced the curve of cases was the use of NPIs. However, we identified that the implementation of lockdowns was a crucial factor that was independent of other factors, as demonstrated in Figure 4, and showed that as the pandemic progressed, the city benefited from the drastic drop in the number of cases. Additionally, in terms of epidemiology, vaccination itself helped decrease cases independently of other strategies. More research will be carried out in the future to gain a more comprehensive understanding.

Autoregressive models have been widely used to predict infectious diseases, especially COVID-19 [60]. SARIMA models were accurate and reliable in predicting deaths in 12 countries that presented high incidences for COVID-19. Importantly, the models properly adapted to the implementation of data, despite seasonality and complex patterns [61]. SARIMA and ARIMA models were used to generate a 60-day forecast of cumulative COVID-19 cases for the top 16 countries. The models were country-specific and were able to classify three categories of case growth [62]; the autoregressive models properly predicted dengue infection [63], monkeypox disease [64], and cases of hemorrhagic fever with renal syndrome [65].

The changes in epidemiological patterns shown in our analysis varied as the pandemic progressed, which is expected in time-series assessments [66]. Thus, the evolution of control measures and the management of cases had to be correlated with the variability of the patterns exhibited in one urban setting. In the case of Araraquara, symptoms, for example, presented different patterns during our analysis. For instance, initially, the community and governors focused on symptoms related to olfactory and taste disorders [67–71], but as shown in Figure 1, it is possible to highlight that the symptoms most related to positive cases were fever, cough, and sore throat (and those described as *Others*). Furthermore, the relative rate of positive and asymptomatic cases showed that the silent spread of the disease once voluntary consent to be tested was provided frequently occurred after the symptoms were experienced. Understanding these pattern changes could also contribute to greater community adherence to undergoing mass testing programs.

The more aspects revealed, the more reasonable the disease perception and understanding were; however, this also demonstrated how complex it is to produce accurate forecasts due to multi-factor deliberations. Non-temporal external static variables could be considered; in other words, a location in a city, for example, is a static variable, because it has multiple inherent features that make it unique.

The Brazilian political, economical, and social scenarios influenced the deployment of public health policies at a time when SARS-CoV-2 spread was a complicated issue with multiple contributing factors [72] also occurred in Araraquara. Another important aspect is the anti-vaccine or anti-public-media movement [73–75], which is frequently fueled by influential misinformation [76]. Aside from these anti-social factors, various environmental and anthropogenic disaster contexts have emerged as being interconnected in the fight against COVID-19 [77].

To improve the comprehension of this problem, once implemented in Araraquara, a massive testing program could assist in counting and demonstrating the value of preventing under-reporting; understanding the origins of epidemics; and the ability to take immediate action to ensure the containment of infected people, whether they are symptomatic or not. The significant efficacy of detecting such profiles via an efficient surveillance system, both quantitatively and qualitatively, as well as case analysis, leads to a better understanding of disease dynamics.

The Brazilian cities and states that implemented lockdowns exhibited positive but varied results, and in general the lockdown decisions were made due to crowding and a lack of ICU beds, not as a measure to prevent disease spread [78,79], as demonstrated in Figures 4 and 7 (Section 3.2). Despite the positive results, not all governments decided to implement this type of measure, which may also have contributed in numbers to their territorial neighbors who implemented lockdowns. Considering these factors, therefore, it is important to analyze the periodicity of case numbers and the mapping of key events, in addition to the need for mass tests. Approaches such as those shown in Figure 8 (Section 3.2) could help with this mapping.

These measures may aid in the reduction of forecasting noise. On the one hand, the model should be highly adaptive; on the other hand, time sequences can be fairly complex or noisy, while others can be modeled simply with naive seasonal predictors and output prediction intervals that reflect the prediction uncertainty. This justifies the mapping of events, as well as the mapping of events inducing crowding (holidays) [34] based on community mobility, and local climate changes. For example, the discussions conducted by Wu et al. [46] and by Aragão et al. [36] showed that weather indices and the spread of COVID-19 were related.

## 5. Conclusions

This work aimed to provide a better understanding of the COVID-19 scenario that occurred in the city of Araraquara during 2020-21, considering the import and export

of reported cases and the main symptoms presented by the population. Furthermore, some intriguing questions arose from Figure 7, showing the two components produced by the t-SNE method, which simplified our understanding of urban mobility in the city. Considering these components, we observed that mobility measures could be important for certain contexts, such as holidays, to avoid a worse scenario. It was also observed that after the implementation of lockdowns, there was a drastic decrease in urban mobility rates, which contributed to the drop in the number of cases of COVID-19 after a few days, but the same behavior did not necessarily apply to other NPI implementations.

Different regression models such as SARIMA and the Transformer model could present different results but similar behaviors, that is, the expectation was an increasing number of COVID-19 cases until the moment the lockdown measure was implemented. Nonetheless, because the virus contagion problem was multi-factorial, this factor (the imposition of the lockdown measure) was only one of the possible variables of interest that could influence the dynamics of COVID-19. It was observed that once a lockdown was implemented, the number of cases decreased but did not reach zero, and that after the restrictive measure was lifted, the number of cases increased again, reaching a similar level to before the lockdown. This could be explained by the extreme contagion rate of COVID-19, which maintains its transmission from person to person even with a low mobility rate within the population. Those who travel were still able to maintain viral transmission. Subsequently, when the restrictive measures were lifted, the virus achieved full activity in the population again. We believe that this behavior will continue until the whole population is vaccinated or the number of infected people who cannot be re-infected becomes high enough.

Aside from producing significant findings indicating that we are on the correct track, our existing data were not definitive, and more efforts are needed to provide a full and final decision on the strengthening of forecasting models. To this end, future studies will need to rank elements, taking into consideration the literature's agreement on the determinants of the COVID-19 infection rate, such as the work presented by Fermo et al. [80] identifying more than 50 possible factors. Additionally, it is possible to observe that the use of the FFT could be essential in the mapping of events over the year, helping decision-making governors fight COVID-19. The findings could help us catch seasonal patterns that repeat over a period of a year. Instead of implementing measures for several repeated outbreaks, this would help in planning and implementing measures at a specific frequency. Without a precisely specified periodicity, Fourier spectrum analysis might be useful for comprehending COVID-19 waves that return with cycles of varied lengths (1 year or less). Finally, using a large number of variables as the input may reduce the noise in predicting the case numbers and make the model independent of these COVID-19 case registries.

Identifying a social event related to crowding and the way the disease manifested in the infected group could aid in future control measures and help identify potential events where an NPI should be used. We would need to compare cities in the same region to understand if these containment measures varied in their ability to decrease incidences, which we intend to evaluate in future work.

**Author Contributions:** All authors have made substantial contributions to the conception or design of the work; the acquisition, analysis, and interpretation of data; the creation of the new software used in the work; the drafting of the work or critical revisions for important intellectual content; and the approval of the version to be published, as described below. Conceptualization: D.P.A. and A.M.; methodology: D.P.A., A.G.d.S.J., A.M., C.D., and L.M.G.G.; software: D.P.A.; formal analysis and investigation: D.P.A., A.M., and C.D.; writing—original draft preparation: D.P.A.; writing—review and editing: D.P.A., A.G.d.S.J., A.M., C.D., and L.M.G.G.; funding acquisition: L.M.G.G.; resources: A.M. and L.M.G.G. In addition, all authors agree to be accountable for all aspects of the work and ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was developed with the partial support of Coordination for the Improvement of Higher Education Personnel (CAPES-Brazil), grant numbers 001 and 88881.506890/2020-01, and the National Research Council (CNPq-Brazil), grant number 311640/2018-4.

**Data Availability Statement:** The code and data used in this research can be found at https://github. com/Natalnet/ncovid-araraquara-paper, accessed on 1 January 2023.

Conflicts of Interest: There are no conflict of interest nor competing interest to declare for this work.

## Abbreviations

----

The following abbreviations are used in this manuscript:

. ....

CMK	Community mobility report
DFT	Discrete Fourier transform
FT	Fourier transform
FFT	Fast Fourier transform
HDR	Health department report
IBGE	Brazilian Institute of Geography and Statistics
KL	Kullback–Leiber divergence
MDR	Municipal daily report
MLP	Multilayer perceptron
NPI	Non-pharmaceutical intervention
RMSE	Root-mean-square error
RT-PCR	Real-time polymerase chain reaction
SARIM	A Seasonal Auto-regressive Integrated Moving Average
T-SNE	t-Distributed Stochastic Neighbor Embedding

# References

- 1. Long, B.; Carius, B.M.; Chavez, S.; Liang, S.Y.; Brady, W.J.; Koyfman, A.; Gottlieb, M. Clinical update on COVID-19 for the emergency clinician: Presentation and evaluation. *Am. J. Emerg. Med.* **2022**, *54*, 46–57. [CrossRef] [PubMed]
- de Souza, W.M.; Buss, L.F.; Candido, D.d.S.; Carrera, J.P.; Li, S.; Zarebski, A.E.; Pereira, R.H.M.; Prete, C.A.; de Souza-Santos, A.A.; Parag, K.V.; et al. Epidemiological and clinical characteristics of the COVID-19 epidemic in Brazil. *Nat. Hum. Behav.* 2020, 4, 856–865. [CrossRef] [PubMed]
- 3. WHO—Coronavirus Disease (COVID-19) Pandemic. Available online: https://www.who.int/emergencies/diseases/novelcoronavirus-2019 (accessed on 30 June 2021).
- 4. Bell, D.; Nicoll, A.; Fukuda, K.; Horby, P.; Monto, A.; Hayden, F.; Wylks, C.; Sanders, L.; Tam, J.V. Non-pharmaceutical interventions for pandemic influenza, national and community measures. *Emerg. Infect. Dis.* **2006**, *12*, 88–94. [PubMed]
- Ferguson, N.; Laydon, D.; Nedjati Gilani, G.; Imai, N.; Ainslie, K.; Baguelin, M.; Bhatia, S.; Boonyasiri, A.; Cucunuba Perez, Z.; Cuomo-Dannenburg, G.; et al. *Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce COVID-19 Mortality and Healthcare* Demand; Imperial College COVID-19 Response Team: London, UK, 2020; Volume 20, pp. 1–20. [CrossRef]
- Flaxman, S. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* 2020, 584, 257–261. [CrossRef] [PubMed]
- Gibbs, H.; Liu, Y.; Abbott, S.; Baffoe-Nyarko, I.; Laryea, D.O.; Akyereko, E.; Kuma-Aboagye, P.; Asante, I.A.; Mitjà, O.; Ampofo, W.; et al. Association between mobility, non-pharmaceutical interventions, and COVID-19 transmission in Ghana: A modelling study using mobile phone data. *PLoS Glob. Public Health* 2022, 2, e0000502. [CrossRef]
- 8. Mohammadi, Z.; Cojocaru, M.G.; Thommes, E.W. Human behaviour, NPI and mobility reduction effects on COVID-19 transmission in different countries of the world. *BMC Public Health* **2022**, *22*, 1594. [CrossRef]
- 9. Snoeijer, B.T.; Burger, M.; Sun, S.; Dobson, R.J.; Folarin, A.A. Measuring the effect of Non-Pharmaceutical Interventions (NPIs) on mobility during the COVID-19 pandemic using global mobility data. *NPJ Digit. Med.* **2021**, *4*, 81. [CrossRef]
- 10. Mahmoudi, J.; Xiong, C. How social distancing, mobility, and preventive policies affect COVID-19 outcomes: Big data-driven evidence from the District of Columbia-Maryland-Virginia (DMV) megaregion. *PLoS ONE* **2022**, *17*, e0263820. [CrossRef]
- Wellenius, G.A.; Vispute, S.; Espinosa, V.; Fabrikant, A.; Tsai, T.C.; Hennessy, J.; Dai, A.; Williams, B.; Gadepalli, K.; Boulanger, A.; et al. Impacts of social distancing policies on mobility and COVID-19 case growth in the US. *Nat. Commun.* 2021, 12, 3118. [CrossRef]
- 12. de Souza Melo, A.; da Penha Sobral, A.I.G.; Marinho, M.L.M.; Duarte, G.B.; Vieira, A.A.; Sobral, M.F.F. The impact of social distancing on COVID-19 infections and deaths. *Trop. Dis. Travel Med. Vaccines* **2021**, *7*, 12 . [CrossRef]
- Song, Q.; Mi, Y.; Ruan, L. Pros and cons factors influence population attitudes toward non-pharmaceutical interventions and vaccination during post–COVID-19. *Public Health* 2022, 211, 88–96. [CrossRef]

- 14. Delaporte, I.; Escobar, J.; Peña, W. The distributional consequences of social distancing on poverty and labour income inequality in Latin America and the Caribbean. *J. Popul. Econ.* **2021**, *34*, 1385–1443. [CrossRef]
- 15. Barnett-Howell, Z.; Watson, O.J.; Mobarak, A.M. The benefits and costs of social distancing in high- and low-income countries. *Trans. R. Soc. Trop. Med. Hyg.* **2021**, *115*, 807–819. [CrossRef]
- Bolaño-Ortiz, T.R.; Camargo-Caicedo, Y.; Puliafito, S.E.; Ruggeri, M.F.; Bolaño-Diaz, S.; Pascual-Flores, R.; Saturno, J.; Ibarra-Espinosa, S.; Mayol-Bracero, O.L.; Torres-Delgado, E.; et al. Spread of SARS-CoV-2 through Latin America and the Caribbean region: A look from its economic conditions, climate and air pollution indicators. *Environ. Res.* 2020, 191, 109938. [CrossRef]
- 17. Verschuur, J.; Koks, E.E.; Hall, J.W. Global economic impacts of COVID-19 lockdown measures stand out in high-frequency shipping data. *PLoS ONE* 2021, *16*, e0248818. [CrossRef]
- Mandel, A.; Veetil, V. The Economic Cost of COVID Lockdowns: An Out-of-Equilibrium Analysis. *Econ. Disasters Clim. Chang.* 2020, 4, 1. [CrossRef]
- 19. Adams-Prassl, A.; Boneva, T.; Golin, M.; Rauh, C. The impact of the coronavirus lockdown on mental health: Evidence from the United States. *Econ. Policy* **2022**, *37*, 139–155. [CrossRef]
- Zhou, F.; Hu, T.J.; Zhang, X.Y.; Lai, K.; Chen, J.H.; Zhou, X.H. The association of intensity and duration of non-pharmacological interventions and implementation of vaccination with COVID-19 infection, death, and excess mortality: Natural experiment in 22 European countries. J. Infect. Public Health 2022, 15, 499–507. [CrossRef]
- Socolovithc, R.L.; Fumis, R.R.L.; Tomazini, B.M.; Pastore, L.; Galas, F.R.B.G.; de Azevedo, L.C.P.; Costa, E.L.V. Epidemiology, outcomes, and the use of intensive care unit resources of critically ill patients diagnosed with COVID-19 in Sao Paulo, Brazil: A cohort study. *PLoS ONE* 2020, 15, e0243269. [CrossRef]
- Noronha, K.V.M.d.S.; Guedes, G.R.; Turra, C.M.; Andrade, M.V.; Botega, L.; Nogueira, D.; Calazans, J.A.; Carvalho, L.; Servo, L.; Ferreira, M.F. The COVID-19 pandemic in Brazil: Analysis of supply and demand of hospital and ICU beds and mechanical ventilators under different scenarios. *Cadernos de Saúde Pública* 2020, *36*, 115320. [CrossRef]
- 23. Malta, M.; Vettore, M.V.; da Silva, C.M.F.P.; Silva, A.B.; Strathdee, S.A. Political neglect of COVID-19 and the public health consequences in Brazil: The high costs of science denial. *EClinicalMedicine* **2021**, *35*, 100878. [CrossRef] [PubMed]
- Koga, N.M.; Palotti, P.L.D.M.; Miranda, P.A.D.; Pontes, M.; Couto, B.G.D.; Luiz, M.; Soares, V. Analytical capacity as a critical condition for responding to COVID-19 in Brazil. *Policy Soc.* 2023, 42, 117–130. [CrossRef]
- 25. de Almeida, L.; Carelli, P.V.; Cavalcanti, N.G.; do Nascimento, J.D.; Felinto, D. Quantifying political influence on COVID-19 fatality in Brazil. *PLoS ONE* 2022, 17, e0264293. [CrossRef] [PubMed]
- 26. Szwarcwald, C.L.; Boccolini, C.S.; da Silva de Almeida, W.; Filho, A.M.S.; Malta, D.C. COVID-19 mortality in Brazil, 2020-21: Consequences of the pandemic inadequate management. *Arch. Public Health* **2022**, *80*, 255. [CrossRef] [PubMed]
- 27. Prado, M.F.D.; Antunes, B.B.D.P.; Bastos, L.D.S.L.; Peres, I.T.; Silva, A.D.A.B.D.; Dantas, L.F.; Baião, F.A.; Maçaira, P.; Hamacher, S.; Bozza, F.A. Analysis of COVID-19 under-reporting in Brazil. *Rev. Bras. Ter. Intensiv.* **2020**, *32*, 224. [CrossRef]
- Touchton, M.; Knaul, F.M.; Arreola-Ornelas, H.; Porteny, T.; Sánchez, M.; Méndez, O.; Faganello, M.; Edelson, V.; Gygi, B.; Hummel, C.; et al. A partisan pandemic: state government public health policies to combat COVID-19 in Brazil. *BMJ Glob. Health* 2021, 6, e005223. [CrossRef]
- 29. Rousseff, D.; Belchior, M.; Pereira, P.E.; Diretor-Executivo, N.; Da, S.; Côrtes, C.; Tai, D.W. Sinopse do Censo Demográfico: 2010. In *Censo 2010*; IBGE: Rio de Janeiro, Brazil, 2011; p. 265.
- Prefeitura Municipal de Araraquara—Lockdown: Araraquara Tem Queda no Numero de Casos e Media Movel por COVID-19. Available online: http://www.araraquara.sp.gov.br/noticias/2021/marco/10/lockdown-araraquara-tem-queda-no-numerode-casos-e-media-movel-por-covid-19 (accessed on 3 July 2021).
- 31. Ferreira, L.S.; Maria, F.; Marquitti, D.; Lopes, R.; Silva, P.D.; Borges, M.E.; Gomes, M.F.D.C.; Cruz, O.G.; Kraenkel, R.A.; Coutinho, R.M.; et al. Estimating the impact of implementation and timing of the COVID-19 vaccination programme in Brazil: A counterfactual analysis. *Lancet Reg. Health Am.* **2023**, *17*, 100397. [CrossRef]
- Whitaker, M.; Elliott, J.; Bodinier, B.; Barclay, W.; Ward, H.; Cooke, G.; Donnelly, C.A.; Chadeau-Hyam, M.; Elliott, P. Variant-specific symptoms of COVID-19 in a study of 1,542,510 adults in England. *Nat. Commun.* 2022, 13, 6856. [CrossRef]
- Auler, A.C.; Cássaro, F.A.; da Silva, V.O.; Pires, L.F. Evidence that high temperatures and intermediate relative humidity might favor the spread of COVID-19 in tropical climate: A case study for the most affected Brazilian cities. *Sci. Total Environ.* 2020, 729, 139090. [CrossRef]
- Aragão, D.P.; Santos, D.H.D.; Mondini, A.; Gonçalves, L.M.G. National Holidays and Social Mobility Behaviors: Alternatives for Forecasting COVID-19 Deaths in Brazil. Int. J. Environ. Res. Public Health 2021, 18, 11595. [CrossRef]
- Tandukar, S.; Khanal, R.; Manandhar, R.; Pandey, A.; Sthapit, N.; Sherchan, S.P.; Haramoto, E. Possibility of Detection of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) through Wastewater in Developing Countries. *Water* 2021, 13, 3412. [CrossRef]
- 36. Aragão, D.P.; Oliveira, E.V.; Bezerra, A.A.; dos Santos, D.H.; da Silva Junior, A.G.; Pereira, I.G.; Piscitelli, P.; Miani, A.; Distante, C.; Cuno, J.S.; et al. Multivariate data driven prediction of COVID-19 dynamics: Towards new results with temperature, humidity and air quality data. *Environ. Res.* 2022, 204, 112348. [CrossRef]
- Liu, Z.; Liang, Q.; Liao, H.; Yang, W.; Lu, C. Effects of short-term and long-term exposure to ambient air pollution and temperature on long recovery duration in COVID-19 patients. *Environ. Res.* 2023, 216, 114781. [CrossRef]

- Jing, N.; Shi, Z.; Hu, Y.; Yuan, J. Cross-sectional analysis and data-driven forecasting of confirmed COVID-19 cases. *Appl. Intell.* 2022, 52, 3303–3318. [CrossRef]
- 39. Mwitondi, K.S.; Said, R.A. A framework for data-driven solutions with covid-19 illustrations. Data Sci. J. 2021, 20, 36. [CrossRef]
- Rusch, T.; Han, Y.; Liang, D.; Hopkins, A.R.; Lawrence, C.V.; Maoz, U.; Paul, L.K.; Stanley, D.A.; Adolphs, R.; Alvarez, R.M.; et al. COVID-Dynamic: A large-scale longitudinal study of socioemotional and behavioral change across the pandemic. *Sci. Data* 2023, 10, 71. [CrossRef]
- COVID-19 Community Mobility Report. Available online: https://www.google.com/covid19/mobility?hl=en | (accessed on 30 June 2021).
- BRASIL.IO—Boletins Informativos e Casos do Coronavírus por Município por dia. Available online: https://brasil.io/dataset/ covid19/caso (accessed on 30 June 2021).
- 43. van der Maaten, L.; Hinton, G.E. Visualizing High-Dimensional Data Using t-SNE. J. Mach. Learn. Res. 2008, 9, 2579–2605.
- 44. Hinton, G.; Roweis, S. Stochastic Neighbor Embedding. Adv. Neural Inf. Process. Syst. 2003, 15, 833–840.
- 45. Jollife, I.T.; Cadima, J. Principal Component Analysis: A Review and Recent Developments. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* 2016, 374, 20150202. [CrossRef]
- 46. Wu, Y.; Jing, W.; Liu, J.; Ma, Q.; Yuan, J.; Wang, Y.; Du, M.; Liu, M. Effects of temperature and humidity on the daily new cases and new deaths of COVID-19 in 166 countries. *Sci. Total Environ.* **2020**, *729*, 139051. [CrossRef]
- 47. Brassey, J.; Heneghan, C.; Mahtani, K.R.; Aronson, J.K. *Do Weather Conditions Influence the Transmission of the Coronavirus* (*SARS-CoV-2*); Technical Report; The Centre for Evidence-Based Medicine: Oxford, UK, 2020.
- Ala'raj, M.; Majdalawieh, M.; Nizamuddin, N. Modeling and forecasting of COVID-19 using a hybrid dynamic model based on SEIRD with ARIMA corrections. *Infect. Dis. Model.* 2021, 6, 98–111. [CrossRef] [PubMed]
- 49. Goic, M.; Bozanic-Leal, M.S.; Badal, M.; Basso, L.J. COVID-19: Short-term forecast of ICU beds in times of crisis. *PLoS ONE* 2021, 16, e0245272. [CrossRef] [PubMed]
- Rguibi, M.A.; Moussa, N.; Madani, A.; Aaroud, A.; Zine-dine, K. Forecasting Covid-19 Transmission with ARIMA and LSTM Techniques in Morocco. SN Comput. Sci. 2022, 3, 133. [CrossRef] [PubMed]
- Siddikur, M.; Id, R.; Hossain, A.; Id, C.; Amrin, M. Accuracy comparison of ARIMA and XGBoost forecasting models in predicting the incidence of COVID-19 in Bangladesh. *PLoS Glob. Public Health* 2022, 2, e0000495. [CrossRef]
- 52. Qi, B.; Liu, N.; Yu, S.; Tan, F.; Qi, B.; Liu, N.; Yu, S.; Tan, F. Comparing COVID-19 Case Prediction Between ARIMA Model and Compartment Model—China, December 2019–April 2020. *China CDC Wkly.* **2022**, *4*, 1185–1188. [CrossRef] [PubMed]
- 53. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* 2017, 30. [CrossRef]
- 54. Wen, Q.; Zhou, T.; Zhang, C.; Chen, W.; Ma, Z.; Yan, J.; Sun, L. Transformers in time series: A survey. arXiv 2022. arXiv:2202.07125.
- 55. Mohd Razali, N.; Bee Wah, Y. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *J. Stat. Model. Anal.* **2011**, *2*, 21–33.
- Song, Q.; Esogbue, A. A New Algorithm for Automated Box-Jenkins ARMA Time Series Modeling Using Residual Autocorrelation/Partial Autocorrelation Functions. *IEMS* 2006, 5, 116–125.
- 57. Anderson, O.D. The Box-Jenkins approach to time series analysis. *RAIRO—Oper. Res. Rech. OpÉrationnelle* **1977**, *11*, 3–29. [CrossRef]
- Johansson, M.A.; Quandelacy, T.M.; Kada, S.; Prasad, P.V.; Steele, M.; Brooks, J.T.; Slayton, R.B.; Biggerstaff, M.; Butler, J.C. SARS-CoV-2 Transmission From People Without COVID-19 Symptoms. JAMA Netw. Open 2021, 4, e2035057. [CrossRef]
- Lauer, S.A.; Grantz, K.H.; Bi, Q.; Jones, F.K.; Zheng, Q.; Meredith, H.R.; Azman, A.S.; Reich, N.G.; Lessler, J. The incubation period of coronavirus disease 2019 (CoVID-19) from publicly reported confirmed cases: Estimation and application. *Ann. Intern. Med.* 2020, 172, 577–582. [CrossRef]
- 60. Yadav, S.K.; Akhter, Y. Statistical Modeling for the Prediction of Infectious Disease Dissemination With Special Reference to COVID-19 Spread. *Front. Public Health* **2021**, *9*, 680. [CrossRef]
- 61. Perone, G. Using the SARIMA Model to Forecast the Fourth Global Wave of Cumulative Deaths from COVID-19: Evidence from 12 Hard-Hit Big Countries. *Econometrics* **2022**, *10*, 18. [CrossRef]
- ArunKumar, K.E.; Kalaga, D.V.; Kumar, C.M.S.; Chilkoor, G.; Kawaji, M.; Brenza, T.M. Forecasting the dynamics of cumulative COVID-19 cases (confirmed, recovered and deaths) for top-16 countries using statistical machine learning models: Auto-Regressive Integrated Moving Average (ARIMA) and Seasonal Auto-Regressive Integrated Moving Average (SARIMA). *Appl. Soft Comput.* 2021, 103, 107161. [CrossRef]
- 63. Martinez, E.Z.; da Silva, E.A.S.; dal Fabbro, A.L. A SARIMA forecasting model to predict the number of cases of dengue in Campinas, State of São Paulo, Brazil. *Rev. Soc. Bras. Med. Trop.* **2011**, *44*, 436–440. [CrossRef]
- Pramanik, A.; Sultana, S.; Rahman, M.S. Time Series Analysis and Forecasting of Monkeypox Disease Using ARIMA and SARIMA Model. In Proceedings of the 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 3–5 October 2022; pp. 1–7. [CrossRef]
- 65. Qi, C.; Zhang, D.; Zhu, Y.; Liu, L.; Li, C.; Wang, Z.; Li, X. SARFIMA model prediction for infectious diseases: Application to hemorrhagic fever with renal syndrome and comparing with SARIMA. *BMC Med. Res. Methodol.* **2020**, *20*, 243 . [CrossRef]
- Wang, W.K.; Chen, I.; Hershkovich, L.; Yang, J.; Shetty, A.; Singh, G.; Jiang, Y.; Kotla, A.; Shang, J.Z.; Yerrabelli, R.; et al. A Systematic Review of Time Series Classification Techniques Used in Biomedical Applications. *Sensors* 2022, 22, 8016. [CrossRef]

- 67. Ercoli, T.; Masala, C.; Pinna, I.; Orofino, G.; Solla, P.; Rocchi, L.; Defazio, G. Qualitative smell/taste disorders as sequelae of acute COVID-19. *Neurol. Sci.* 2021, *42*, 4921–4926. [CrossRef]
- 68. Dell'Era, V.; Farri, F.; Garzaro, G.; Gatto, M.; Valletti, P.A.; Garzaro, M. Smell and taste disorders during COVID-19 outbreak: Cross-sectional study on 355 patients. *Head Neck* **2020**, *42*, 1591–1596. [CrossRef]
- Tan, B.K.J.; Han, R.; Zhao, J.J.; Tan, N.K.W.; Quah, E.S.H.; Tan, C.J.W.; Chan, Y.H.; Teo, N.W.Y.; Charn, T.C.; See, A.; et al. Prognosis and persistence of smell and taste dysfunction in patients with COVID-19: Meta-analysis with parametric cure modelling of recovery curves. *BMJ* 2022, *378*, e069503. [CrossRef]
- 70. Boscolo-Rizzo, P.; Polesel, J.; Vaira, L.A. Smell and taste dysfunction after covid-19. BMJ 2022, 378, o1653. [CrossRef] [PubMed]
- da Costa, K.V.; Carnaúba, A.T.L.; Rocha, K.W.; de Andrade, K.C.L.; Ferreira, S.M.; de L. Menezes, P. Olfactory and taste disorders in COVID-19: a systematic review. *Braz. J. Otorhinolaryngol.* 2020, *86*, 781–792. [CrossRef] [PubMed]
- 72. Xavier, D.R.; e Silva, E.L.; Lara, F.A.; e Silva, G.R.; Oliveira, M.F.; Gurgel, H.; Barcellos, C. Involvement of political and socioeconomic factors in the spatial and temporal dynamics of COVID-19 outcomes in Brazil: A population-based study. *Lancet Reg. Health Am.* **2022**, *10*, 100221. [CrossRef]
- Khadafi, R.; Nurmandi, A.; Qodir, Z.; Misran. Hashtag as a new weapon to resist the COVID-19 vaccination policy: A qualitative study of the anti-vaccine movement in Brazil, USA, and Indonesia. *Hum. Vaccines Immunother.* 2022, 18, 2042135. [CrossRef] [PubMed]
- 74. Fernandez, M.; Matta, G.; Paiva, E. COVID-19, vaccine hesitancy and child vaccination: Challenges from Brazil. *Lancet Reg. Health Am.* **2022**, *8*, 100246. [CrossRef]
- 75. Germani, F.; Biller-Andorno, N. The anti-vaccination infodemic on social media: A behavioral analysis. *PLoS ONE* **2021**, *16*, e0247642. [CrossRef]
- 76. De Oliveira, I.S.; Cardoso, L.S.; Ferreira, I.G.; Alexandre-Silva, G.M.; de Cássia da Silva Jacob, B.; Cerni, F.A.; Monteiro, W.M.; Zottich, U.; Pucca, M.B. Anti-vaccination movements in the world and in Brazil. *Rev. Soc. Bras. Med. Trop.* 2022, 55, 2022. [CrossRef]
- 77. Shekhar, H.; Rautela, M.; Maqsood, M.; Paris, R.; de León, R.M.F.; Romero-Aguirre, M.F.; Balinos, M.; Velázquez, M.E.; Amri, G.S.; Rahman, T.; et al. Are leading urban centers predisposed to global risks- An analysis of the global south from COVID-19 perspective. *Habitat Int.* 2022, 121, 102517. [CrossRef]
- 78. de Moraes, R.F.; Russell, L.B.; da Silva, L.L.S.; Toscano, C.M. Effects of non-pharmaceutical interventions on social distancing during the COVID-19 pandemic: Evidence from the 27 Brazilian states. *PLoS ONE* **2022**, *17*, e0265346. [CrossRef]
- Runge, M.; Richardson, R.A.K.; Clay, P.A.; Bell, A.; Holden, T.M.; Singam, M.; Tsuboyama, N.; Arevalo, P.; Fornoff, J.; Patrick, S.; et al. Modeling robust COVID-19 intensive care unit occupancy thresholds for imposing mitigation to prevent exceeding capacities. *PLoS Glob. Public Health* 2022, 2, e0000308. [CrossRef]
- Fermo, P.; Artíñano, B.; De Gennaro, G.; Pantaleo, A.M.; Parente, A.; Battaglia, F.; Colicino, E.; Di Tanna, G.; Goncalves da Silva Junior, A.; Pereira, I.G.; et al. Improving indoor air quality through an air purifier able to reduce aerosol particulate matter (PM) and volatile organic compounds (VOCs): Experimental results. *Environ. Res.* 2021, 197, 111131. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.