

Borough-level COVID-19 forecasting in London using deep learning techniques and a novel MSE-Moran's I loss function

Frederik Olsen^{a,*}, Calogero Schillaci^b, Mohamed Ibrahim^a, Aldo Lipani^a

^a Department of Civil, Environmental and Geomatic Engineering, University College London (UCL), England

^b Department of Agricultural and Environmental Science, University of Milan, Via Celoria 2 Milan, Italy

ARTICLE INFO

Keywords:

COVID-19
Deep Learning
LSTM
Epidemiological Modelling
Pandemic

ABSTRACT

Following its identification in late 2019, COVID-19 has spread around the globe, and been declared a pandemic. With this in mind, modelling the spread of COVID-19 remains important for responding effectively. To date research has focused primarily on modelling the spread of COVID-19 on national and regional scales with just a few studies doing so on a city and sub-city scale. However, no attempts have yet been made to design and optimize a model explicitly for accurately forecasting the spread of COVID-19 at sub-city scale. This research aimed to address this research gap by developing an experimental LSTM-ANN deep learning model. The model is largely autoregressive in nature as it considers temporally lagged borough-level COVID-19 cases data from the last 9 days, but also considers temporally lagged (i) borough-level NO₂ concentration data, (ii) government stringency data, and (iii) climatic data from the last 9 days, as well as non-temporally variable borough-level urban characteristics data when modelling and forecasting the spread of the disease. The model was also encouraged to learn the spatial relationships between boroughs with regards to the spread of COVID-19 by a novel MSE-Moran's I loss function. Overall, the model's performance appears promising and so the model represents a useful tool for assisting the decision making and interventions of governing bodies within cities. A sensitivity analysis also indicated that of the non COVID-19 variables, the government stringency is particularly important in the modelling process, with this being closely followed by the climatic variables, the NO₂ concentration data, and finally the urban characteristics data. Additionally, the introduction of the novel MSE-Moran's I loss function appeared to improve the model's forecasting performance, and so this research has implications at the intersection of deep learning and disease modelling. It may also have implications within spatio-temporal forecasting more generally because such a feature may have the potential to improve forecasting in other spatio-temporal applications

Introduction

Alongside continuing advancements in healthcare and epidemiology, the management, understanding, modelling, and prediction of disease within human society has remained prominent within literature in recent years (e.g Mao & Bian, 2010; Mei et al., 2015; Erraguntla et al., 2019; Scarpino & Petri, 2019; Feng & Jiao, 2021). At present the importance of this literature is exceptionally apparent with a new β -coronavirus called SARS-CoV-2 (Lake, 2020) having spread a respiratory disease called COVID-19 around the world, infecting 77,864,273 and killing 1,712,072 as of 22nd December 2020 (Worldometer, 2020). The disease received pandemic status (WHO, 2020) and presents a considerable challenge to human society, and so requires that societies and governments worldwide respond effectively.

The academic community has responded to the threat presented by COVID-19 with numerous forecasting studies. These studies vary in scope and purpose, with some aiming to accurately forecast future values in order to highlight where and when COVID-19 will be most prevalent, whilst others aim to analyse the effectiveness of responses, and others to assess the impact of future measures. These forecasting studies operate on various different spatial scales. Of these scales, perhaps the most extensively studied is the country scale. Examples of this include Chimmula & Zhang (2020), Parbat & Chakraborty (2020), Shastri et al. (2020), Balli (2020), Wang et al. (2020), Yonar et al. (2020), Fanelli & Piazza (2020) and Semenova et al. (2020) as well as preprints Dehesh et al. (2020), De Castro (2020), Toda (2020) and Ibrahim et al. (2020). Also covered is the regional scale within countries, with examples including Ribeiro et al. (2020), Roosa et al. (2020), Yang

* Corresponding author.

et al. (2020) and Melo et al. (2020). Below the regional scale, a few studies have made forecasts on the city scale. Examples of this include Ramasamy et al. (2020) as well as preprints Sugishita et al. (2020) and Yan et al. (2020).

On an even smaller scale such as within cities, literature is even more sparse. The only study found is Goscé et al. (2020) which studied the impact of the early lockdown measures at the beginning of the pandemic and then calibrated a simple model to consider the impact of different scenarios in London. However, Goscé et al. (2020) made no attempt to evaluate this fitted model’s predictive ability by comparing its forecasted future values against unseen observed values. Furthermore, Goscé et al. (2020) made use of a simple model designed to evaluate the impact of different scenarios and so it would likely not perform well with this purpose in mind. Therefore, there remains a research gap for a study which aims to develop a novel forecasting model designed and optimized explicitly to provide accurate borough-level forecasts within a city.

The poor coverage of research into this area is perhaps somewhat surprising considering that more than 4 billion people worldwide live in urban areas (Ritchie, 2018). Furthermore, cities often exhibit high population densities and high internal mobility which can result in frequent intimate contact between individuals (Yang et al., 2008). As noted by Mao & Bian (2010), cities and urban areas therefore often play an important role in fostering the transmission of infectious diseases. Additionally, the dramatic inequality often observed within cities related to variables such as wealth, population density and deprivation (Lee, 2019) may lend itself to a spatially heterogenous response between boroughs/districts. A geospatial approach to managing COVID-19 within cities could therefore be important for improving public health by identifying where and when COVID-19 will be most prevalent within an already vulnerable environment and enabling the spatial units that require intervention and assistance to receive it.

With a research gap identified, it is important to consider models within infectious disease epidemiology literature to address this gap. Perhaps the most popular is the Susceptible-Infected-Recovered (SIR) model, originally proposed by Kermack and McKendrick (1927), which is an epidemiological compartmental model that divides the population into three categories: (i) susceptible, (ii) infected, and (iii) recovered. By considering the rate at which individuals move between them over time, the proportion of individuals belonging to each category within a population can be computed (Johnson and McQuarrie, 2009). This can then

be applied to simulate system behaviours and forecast disease epidemiology over time. Various adaptations of the SIR model have already been applied by the likes of Yang et al. (2020) and Sugishita et al. (2020) to provide forecasts specifically of COVID-19 data. The main advantage of the SIR model is that it is not computationally intensive. However, the model requires many assumptions concerning the rates of movements between categories and the initial structure of the system, with the model only being valid under these assumptions (Ibrahim et al., 2020). For this reason, the SIR model is rejected.

Statistical models explicitly designed for time series forecasting can also be used. The most popular are the (i) Autoregressive-Integrated-Moving-Average (ARIMA) and (ii) Exponential Smoothing (ES) models (Hyndman & Athanasopoulos, 2018) which make linear assumptions over past values to predict those in the future. The ARIMA model has already been applied by the likes of Ribeiro et al. (2020), Yonar et al. (2020) and Dehesh et al. (2020) whilst the ES model has been applied by the likes of Petropoulos & Makridakis (2020) and Yonar et al. (2020) to forecast COVID-19 data. Advantages of both the ARIMA and ES models include not being computationally intensive and not requiring large datasets. They are also particularly effective at forecasting a time series with linear properties as they make linear assumptions. However, they often struggle to capture non-linear components within data (Abbasi-mehr et al., 2020). Such a criticism is also noted by Chimmula & Zhang (2020), which states that since infectious disease time series data displays complex non-linear patterns, models that make linear assumptions may not be appropriate for forecasting COVID-19 data. As such these models are not favored.

Machine Learning (ML) is a subsection of computer science and a branch of artificial intelligence (AI) concerned with the construction and study of systems that can learn from datasets without explicit programming (Voyant et al., 2017). ML models can be deployed to forecast a time series, with the most prevalent being based upon (i) Support Vector Machines (SVMs) and (ii) Artificial Neural Networks (ANNs). SVMs are based upon statistical learning theory (Vapnik, 2013) and are used for solving classification and regression tasks (Han et al., 2011). SVMs aim to determine a hyperplane that can separate classes of data points with a maximal margin before mapping data to higher dimensional feature space with kernel functions where classes are not linearly separable (Wu et al., 2004; Han et al., 2011; Parmezan et al., 2019). SVMs can be applied to time series forecasting using Support Vector Regression (SVR) (Wu et al., 2004). ANNs refer to a set of methods which

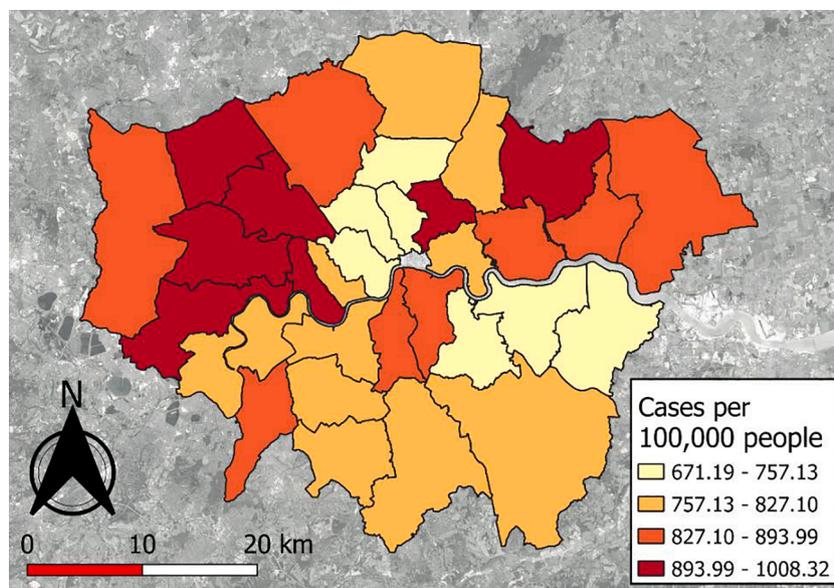


Figure 1. Showing the total confirmed COVID-19 cases per 100,000 people of each borough within London as of 2020-10-16. Data sourced from GLA (2020a) and GLA (2020b).

process input signals and transform them into outputs (Wesolowski & Suchacz, 2012) using collections of connected artificial neurons in a similar way to animal brains (Haykin, 2010). ANNs are diverse in terms of architectures and applications, with the likes of Ghosh & Guha (2010) using a simple Multilayer Perceptron (MLP) network, and Ture & Kurt (2006) using both an MLP and Time Delay Neural Network to predict viral infections. However, at present state-of-the-art Recurrent Neural Networks (RNNs) are often favored for time series applications as they feature recursive, self-reflective connections which provide the network with a ‘memory’ and enable them to effectively capture long-term temporal dependencies and variable length observations (Che et al., 2018). Of the RNNs available, Long Short-term Memory (LSTM) networks are especially popular and effective (Greff et al., 2016), largely due to their solution to the exploding/vanishing gradient problems encountered with other RNNs (Parmezan et al., 2019). SVR has been applied by the likes of Parbat & Chakraborty (2020), Ballı (2020) and Ribeiro et al. (2020), whilst LSTM networks have been applied by the likes of Chimmula & Zhang (2020), Shastri et al. (2020), Wang et al. (2020) and Yan et al. (2020) as well as Ibrahim et al. (2020) which applied a Variational LSTM-Autoencoder model to forecasts COVID-19 case data. Additionally, Convolutional Neural Networks have been successfully used in environmental sciences (James et al., 2021), fire susceptibility modelling (Anderson-Bell et al., 2021), sustainability issues (Kizilcec et al., 2022) and networking (Schelbourne et al., 2022). The main advantage of both ML methods over others presented is their ability to capture non-linear data components which makes them appropriate for forecasting disease epidemiology (Chimmula & Zhang, 2020). Therefore, despite being computationally intensive and often requiring larger datasets, a machine learning approach is favored. Of the two methods, ANNs are favored over SVMs because they offer greater flexibility due to the possibility of model architecture combination/hybridisation and universal approximation (Abbasimehr et al., 2020).

The aim of this research is therefore to develop a novel COVID-19 forecasting model capable of modelling the spread of COVID-19 at borough-level within a city using machine learning methods with the intention of accurately predicting future values. The following intermediate objectives have been identified in order to support this:

- 1) Design a forecasting model capable of making borough-level forecasts within a city multiple time-steps into the future.
- 2) Apply the model to make forecasts using unseen data.
- 3) Evaluate the performance of the model.

Materials and methods

Study Area

This experiment is conducted within the city of London primarily because borough-level COVID-19 data is available. Also, London displays a spatially heterogeneous impact (see [Figure 1](#), with boroughs behaving differently from others within the city. For example, boroughs such as Greenwich and Islington display comparatively low COVID-19 cases per 100,000 (671.19 and 676.47 respectively) compared to the likes of Harrow (1008.32) and Ealing (1001.99). London therefore provides a suitable study area for testing a borough-level forecasting system that aims to model the individual way that each borough behaves.

Model assumptions

The model makes 6 key assumptions in order to provide COVID-19 forecasts. The first is that features can be extracted from historical COVID-19 data in order to predict future values in an autoregressive context. This assumption is informed by the identification of statistically significant temporal autocorrelation and statistically significant temporal partial autocorrelation in the historical COVID-19 data.

The second assumption is that the intensity of COVID-19 within a given borough is likely to be similar to others that are located nearby due to the spatially constituted nature of infectious disease epidemiology, and as such the spatial relationships between boroughs with regards to COVID-19 intensity should be extracted when predicting future values. This assumption was also informed by the identification of statistically significant space-time autocorrelation and statistically significant space-time partial autocorrelation within the COVID-19 data, indicating that borough-wise COVID-19 new case values are correlated with temporally lagged values at geometrically adjacent boroughs.

The third assumption is that the spread of COVID-19 is affected by mobility within a city, and so borough-wise historical mobility data should be considered when making predictions as it is hypothesized that during periods of higher mobility people are more likely to come into close contact with others and facilitate the spread of COVID-19. Such an assumption is informed by Carteni et al. (2020) which identified a relationship between mobility and the spread of COVID-19.

The fourth assumption is that the spread of COVID-19 is affected by climatic features such as precipitation and temperature, and so historical climatic data can be considered when making predictions. This assumption is informed by the likes of Ahmadi et al. (2020) identifying a negative relationship between humidity and the spread of COVID-19 and Carteni et al. (2020) a negative relationship between temperature and COVID-19. However, it is also hypothesized that climate acts as a proxy for mobility, whereby during periods of low precipitation and/or higher temperatures people are more likely to leave their homes and come into close contact with others, thus facilitating the spread of COVID-19.

The fifth assumption is that the measures taken by governing bodies over time affects the spread of COVID-19, and so the historical government response measures should be considered when making borough-wise COVID-19 predictions. This is informed by the likes of Achuo (2020) and Kim & Castro (2020) which identified that stricter government responses to contain the spread of COVID-19 were successful in doing so.

The sixth and last assumption is that the urban/demographic features of each borough should be considered when making predictions. This assumption is informed by the likes of Ahmadi et al. (2020), Ibrahim et al. (2020), Guan et al. (2020), Ho et al. (2020), Prats-Urbe et al. (2020) which identified relationships between the spread of COVID-19 and variables such as population density, age and ethnic backgrounds.

Data types and preprocessing

This research made use of temporal data, urban characteristics data and geographic data (shapefile). Temporal data refers to that which changes over time (time series), and a total of four different classes of temporal data were used, all of which were of daily temporal resolution and cover the 248 days between 2020-02-11 and 2020-10-16. Non-temporal features representing the urban characteristics of each borough were also used as it was expected that they may contribute to the dynamics of the spread of COVID-19 amongst boroughs. The geographic data was used for geometric operations and was obtained from Greater London Authority (GLA, 2020c) which contained GIS boundaries of London boroughs.

Temporal data

The first temporal dataset is the borough-level COVID-19 historical data provided by Public Health England (although downloaded from GLA(2020a)) which contains the daily new and total confirmed cases of each London borough.

The second temporal dataset is the climatic data provided by National Centers for Environmental Information (NOAA, 2020), which contains both precipitation (inches) and temperature (F) collected daily at Heathrow.

The third dataset is the government response dataset provided by the

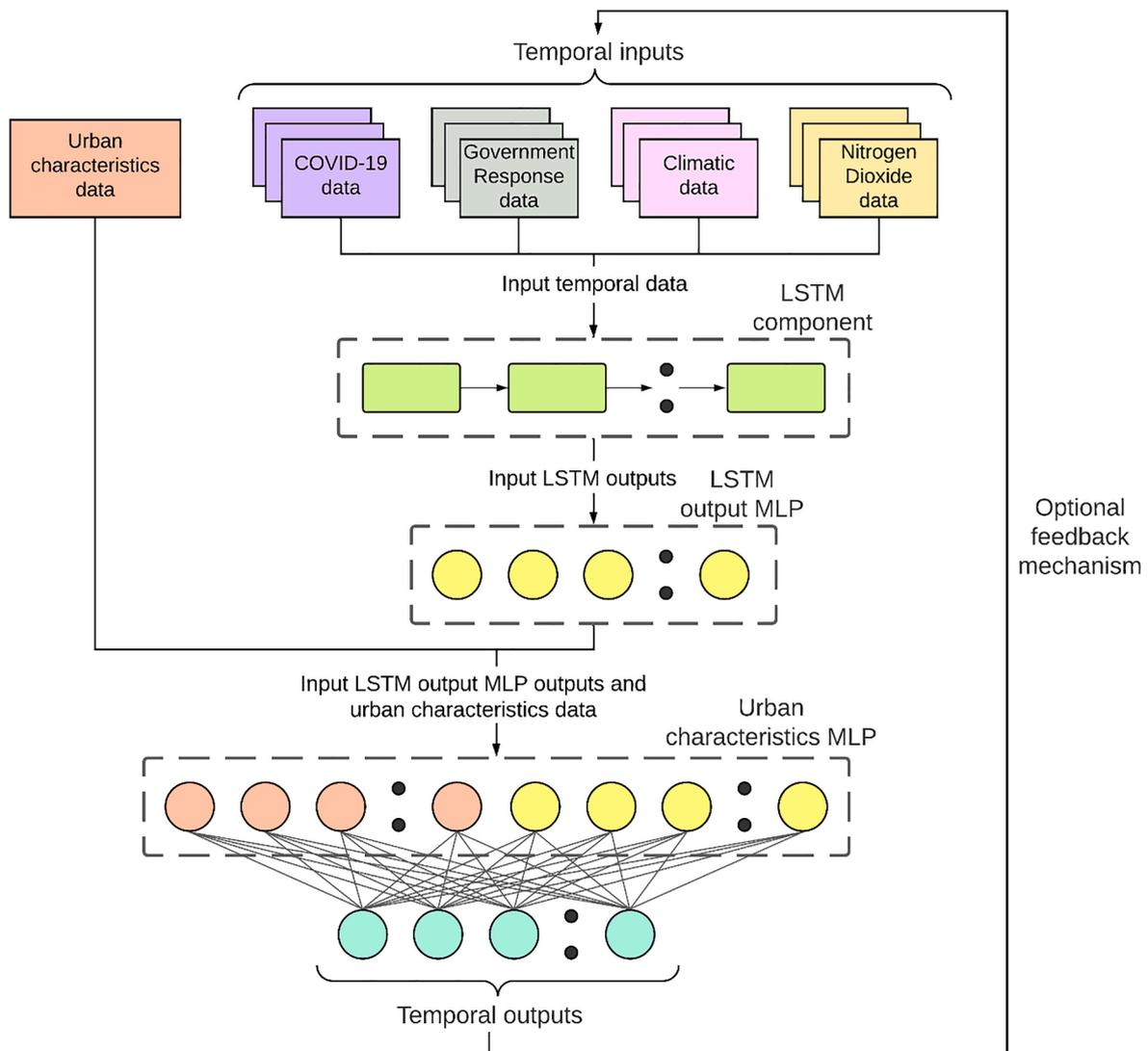


Figure 2. A conceptualisation of the proposed model

Oxford COVID-19 Government Response Tracker (Hale et al., 2020) which collects numerous temporally variable indicators which quantify the stringency/effectiveness of a national government’s response to COVID-19. Specifically, this research made use of the ‘stringency index’ which considers (i) school closures, (ii) workplace closure, (iii) cancellation of public events, (iv) gathering restrictions, (v) public transport closure, (vi) staying at home requirements, (vii) internal movement restrictions, (viii) international travel restrictions, and (ix) public information campaigns.

The final temporal dataset is that of Nitrogen Dioxide (NO₂) concentrations, which acts as a proxy for mobility whereby decreases in NO₂ concentrations may indicate decreases in mobility and vice versa. This dataset is provided by London Air (LAQN, 2020), and contains daily NO₂ concentrations (ppb) from meters collected throughout London, and so represents borough-level recordings.

Urban characteristics data

The features used were (i) Population, (ii) Population density, (iii) Percentage of population aged 65 and over, (iv) Percentage of population from BAME (Black, Asian, Minority Ethnic) backgrounds, (v) Cars per person, (vi) Persons per dwelling, and (vii) Supermarkets per 100,000 people. Features (i)-(iv) were downloaded from Greater London Authority (GLA, 2020b) whilst features (v), (vi) and (vii) were downloaded from Department for Transport (DT, 2020), Ministry of Housing,

Communities & Local Government (MHCLG, 2020) and Pope (2017) respectively.

The selection of features (ii), (iii) and (iv) were informed by literature, with both Ahmadi et al. (2020) and Ibrahim et al. (2020) identifying positive relationships between population density and the spread of COVID-19 whilst Ho et al. (2020) observed that older adults are more likely to test positive for COVID-19, and both Ho et al. (2020) and Prats-Urbe et al. (2020) observed that those of BAME backgrounds within the UK are more likely to test positive.

Meanwhile features (i), (v), (vi) and (vii) are selected with a hypothesized relationship. In the case of (i), it is hypothesized that boroughs with larger populations will exhibit more COVID-19 cases because there are more people who are at risk of contracting the virus. In the case of feature (v), it was initially hypothesized that a negative correlation would exist due to boroughs with more cars per person being wealthier and individuals with access to more cars being less likely to use public transport. For feature (vi), it was hypothesized that households with more members would have a higher likelihood of a household member coming into contact with another infected individual due to more people coming and going from the household, and so boroughs with more people per dwelling would more likely encounter COVID-19. Finally, in the case of variable (vii) it was hypothesized that fewer supermarkets per 100,000 people may lead to more frequent contact between individuals because more people would be required to converge on one

supermarket to shop, with the supermarket acting as a 'hub' for more people. In this way, fewer supermarkets per 100,000 people in a borough would likely facilitate the spread of COVID-19.

Data extraction and feature engineering

Temporal data was extracted between the dates 2020-02-11 and 2020-10-16 for each feature. In the case of the borough-level COVID-19 data, the 'new case' data was used for modelling due to better results and because the model otherwise had a tendency to occasionally predict decreases in total case values, which is impossible given what total cases represent. However, the total case forecasts were converted back into total case forecasts for evaluation. Furthermore, the NO₂ data was not entirely complete, with some boroughs missing time series values, and some boroughs not featuring working stations. In the case of missing values, they were mathematically imputed using linear interpolation. For boroughs without working stations, values at each time-step were calculated as the averages of the boroughs which geometrically shared a border as adjacent boroughs would be expected to exhibit similar NO₂ concentrations. The government stringency and climatic data did not require any adjustments to be made.

Regarding the urban characteristics data, the features Population, Population density, Percentage of population aged 65 and over, Percentage of population from BAME (Black, Asian, Minority Ethnic) backgrounds and Persons per dwelling required no adjustments or alterations to be made and were systematically extracted in their raw formats. However, the Car data from DT (2020) in its raw form represented absolute values and so following extraction the values were divided by Population to produce values which represented Cars per Person. The Supermarkets per 100,000 people feature required more dramatic feature engineering, with the raw data representing point data. In this case, the number of supermarkets located in each borough were calculated geometrically. With the supermarkets per borough counted, the values were divided by respective borough populations and multiplied by 100,000 to yield a representation of Supermarkets per 100,000 people.

Data scaling

Data scaling is required to accelerate the calculations involved in the ML algorithms (Thara et al., 2019). The COVID-19 data was scaled logarithmically, although the presence of zero-values meant that the logarithm of each value plus one was used, since the logarithm of 0 is mathematically impossible (equation 1). Meanwhile, all other data was scaled with a min-max scaler within the range of 0-1 (equation 2).

$$x_s = \log(x + 1) \#(1)$$

$$x_s = \frac{x - \min(x)}{\max(x) - \min(x)} \#(2)$$

Where:

- x_s is the scaled value
- x is the original value

Model Architecture

An LSTM-ANN hybrid model architecture is proposed, which has three main components which transform input data into outputs (see Figure 2). The components are (i) the LSTM component, (ii) the LSTM output MLP component, and (iii) the urban characteristics MLP component. Hyperparameters were selected based upon what gave the optimal performance on a validation set.

LSTM component

The LSTM component represents the principal component of the model and aims to capture the long-term temporal dependencies both within each temporal feature, and between different temporal features.

The LSTM component itself is an RNN, and its main features are (i) a memory cell, which maintains its state over time, and (ii) non-linear gating units which regulate the flow of information into and out of the cell (Greff et al., 2016).

The LSTM forward learning can be summarized as follows:

$$a(t_i) = \sigma(w_a x(t_i) + w_{ha} h(t_i - 1) + b_a) \#(3)$$

$$f(t_i) = \sigma(w_f x(t_i) + w_{hf} h(t_i - 1) + b_f) \#(4)$$

$$c(t_i) = f_i \times c(t_i - 1) + a_i \times \tanh(w_c x(t_i) + w_{hc} (h(t_i - 1) + b_c)) \#(5)$$

$$o(t_i) = \sigma(w_o x(t_i) + w_{ho} h(t_i - 1) + b_o) \#(6)$$

$$h(t_i) = o(t_i) \times \tanh(c(t_i)) \#(7)$$

Where:

- $x(t_i)$ is the input value.
- $h(t_i - 1)$ and $h(t_i)$ are the output values at times $t_i - 1$ and t_i respectively.
- $c(t_i - 1)$ and $c(t_i)$ are the cell states at times $t_i - 1$ and t_i respectively.
- $b = \{b_a, b_f, b_c, b_o\}$ are the biases of the input, forget, internal and output gate.
- $\vec{W}_1 = \{w_a, w_f, w_c, w_o\}$ are the weight matrices of the input gate, forget gate, internal state and output gate.
- $\vec{W}_2 = \{w_{ha}, w_{hf}, w_{hc}, w_{ho}\}$ are recurrent weights.
- $\vec{a} = \{a(t_i), f(t_i), c(t_i), o(t_i)\}$ are the output results for the input gate, forget gate, internal state and output gate.
- Σ and \tanh are activation functions.
- \times represents point-wise multiplication.

And so using the notations outlined above, the LSTM operates as (i) the forget gate $f(t_i)$ receives inputs of $x(t_i)$ and $h(t_i - 1)$ and uses a sigmoid activation to compute the information stored in $c(t_i - 1)$, (ii) the input gate $a(t_i)$ receives $x(t_i)$ and $h(t_i - 1)$ to compute $c(t_i)$, and (iii) the output gate $o(t_i)$ regulates the output from the LSTM cell by considering $c(t_i)$ and applying sigmoid and tanh layers.

The LSTM component of this model featured a single layered LSTM which received inputs of size 67, corresponding to the 67 temporal features (32 x borough-level COVID-19 features, 32 x borough-level NO₂ features, 2 x climatic features, 1 x government response feature), along with the 9 previous timestamps (sequence length selected based upon the performance upon a validation set) of each of the 67 features. This LSTM also had a hidden layer size of 600 and so a corresponding output size of 600.

LSTM output MLP component

The LSTM output MLP component is an MLP and so is formed of perceptrons, which can each be summarized as follows:

$$net = \sum_{i=1}^l w_i x_i + b \#(8)$$

$$y = f(net) \#(9)$$

Where:

- The single neuron comprises l data inputs $x_i \in X$.
- The i^{th} element of X is associated with a synaptic weight w_i , which can assume a negative or positive value reflecting the importance of the input.
- b refers to a 'bias'.
- net refers to the net value.
- $f()$ refers to an activation function
- y refers to the output

And so a perceptron computes net as the result of the linear combi-

nation of inputs with weights and the addition of b , which is the passed to an activation function which introduces non-linear properties to produce output y .

The LSTM output MLP component received the 600 outputs from the LSTM component and transformed these signals into 67 output signals via a hidden layer of size 1200 with ReLU activation, thus extracting features from the outputs of the LSTM component.

Urban characteristics MLP component

The urban feature MLP component was also composed of layers of perceptrons like that outlined in equations 8 and 9. The input signals to this component were (i) the 67 output signals of the LSTM output MLP component, and (ii) the 7 urban features of each of the 32 boroughs, which was flattened to form 224 signals, thus totaling to an input size of 291. These inputs were transformed into 67 output signals via two hidden layers of size 800 and 290, both with ReLU activation. This component thus enabled the model to consider the urban characteristics features associated with each borough and the individual way each borough behaves with respect to each feature alongside the output signals of the LSTM output MLP component and modify the predictions appropriately to provide outputs of all temporal features.

Feedback mechanism

Each forward pass through the components outlined outputs a prediction of all temporal variables which correspond to the next time-step which is one day ahead (single-step) of the input sequence. However, this alone does not enable forecasts more than one step ahead to be made. Therefore, a feedback mechanism was also developed which concatenates the outputs of a forward pass onto the original input sequence whilst removing the most temporally distant existing input from the sequence and feeds the new sequence back into the model. This enables the model to also make predictions that correspond to time-steps which are multiple days (multi-step) into the future.

Model Training and the MSE-Moran's I loss function

The model was trained by feeding batches of inputs to the model, calculating loss, backpropagating error, and updating parameters. This enabled the model to 'learn' from the training data. This was executed for a total of 80 epochs whilst making use of Adam's optimizer at a learning rate of 0.001. However, the model's training process also featured a novel MSE-Moran's I loss function. This function aimed to use Moran's I computations alongside standard Mean Squared Error (MSE) to further encourage the model to more correctly learn the spatial relationships and spatial autocorrelation structures in line with the second model assumption.

MSE is a common loss function used in regression and time series forecasting which represents the sum of squared distances between a target variable and predicted values. The MSE calculation can be summarized as follows:

$$MSE = \frac{\sum_{t=1}^n (Y_t - \bar{Y}_t)^2}{n} \#(10)$$

Where:

- Y_t is the actual value of a point for time period t
- \bar{Y}_t is the predicted value for time period t
- n is the total number of fitted points

Moran's I is a correlation coefficient that measures the overall spatial autocorrelation within a dataset. Similar to correlation coefficients it has values between -1, which represents perfect clustering of dissimilar values and +1 which represents perfect clustering of similar values. Moran's I is calculated as follows:

$$I = \frac{N}{S} \times \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \#(11)$$

Where:

- N is the number of spatial units indexed by i and j
- x is the variable of interest
- \bar{x} is the mean of x
- w_{ij} is a spatial weight matrix
- S is the sum of all w_{ij}

By computing the Moran's I correlation coefficient for the model predictions and comparing the coefficient to that observed within the training data an indication of the similarity in spatial autocorrelation between the two can be observed. Specifically, this was undertaken by computing the Moran's I coefficient for each index along the batch dimension, and averaging to provide a mean Moran's I coefficient amongst the batch of COVID-19 predictions. This value was then subtracted from the mean Moran's I correlation coefficient observed throughout the training data and squared to produce a 'Moran's I loss', and finally multiplied by an adjustable coefficient θ (hyperparameter) as follows:

$$MI_{loss} = \theta \times (MI_{pred} - MI_{obs})^2 \#(12)$$

Where:

- MI_{loss} represents the Moran's I loss.
- θ represents an adjustable coefficient, which represents a hyperparameter.
- MI_{pred} represents the mean Moran's I value calculated along the batch dimension using equation 11 of the model's outputs.
- MI_{obs} represents the mean Moran's I value calculated using equation 11 across the training data.

Therefore, the similarity in spatial autocorrelation between the batch of predictions and that observed across the actual data is computed and weighted by the θ coefficient to produce a Moran's I loss, with a greater loss representing a greater difference between the spatial autocorrelation of the predictions and actual data. A θ value of 0.003 was used in this experiment because it yielded the best results on a validation set. However, prior to performing the calculation in equations 11 and 12 it is worth noting that the model's borough-wise predictions were first reversed transformed, divided by their respective borough populations and multiplied by 100,000 to represent the predicted borough-wise daily new cases per 100,000 people. The decision to convert to the rate (cases per 100,000 people) rather than using the actual values was made because different boroughs have different populations (often considerably). In this way we would expect adjacent boroughs to exhibit greater similarity in rate than absolute values which themselves are also likely to be proportional to borough population as well as intensity of COVID-19 at a given time and so are a poorer representation of COVID-19 intensity. Naturally, this was then also compared to the mean daily cases per 100,000 people Moran's I coefficient observed across the training data (0.0996) rather than that for the actual value data.

Finally, with both the MSE (MSE) and Moran's I loss (MI_{loss}) calculated, the two were added together to produce a total loss (L_{total}):

$$L_{total} = MSE + MI_{loss} \#(13)$$

Therefore, the total loss represents both the distances between the target values and predictions (MSE), and also the difference between the Moran's I coefficient of the predicted cases per 100,000 people and that observed within the training data, the contribution of which is weighted by θ (MI_{loss}).

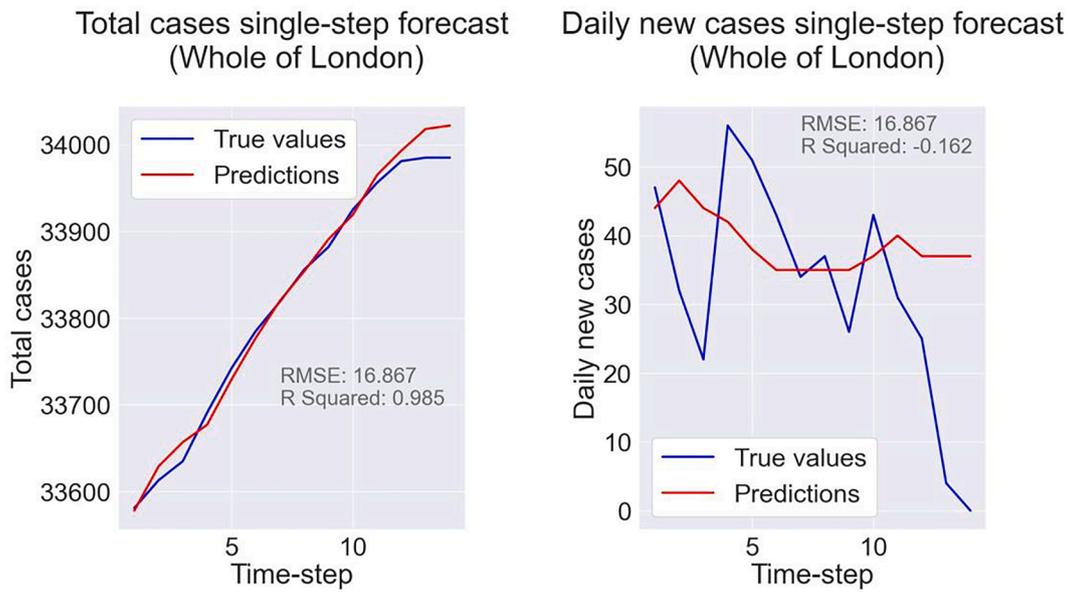


Figure 3. A lineplot showing the total cases and daily new cases single-step forecasts for the whole of London.

Model Evaluation

The model was trained and applied to forecast the next 14 days, with these predictions then being converted into total case forecasts for the purpose of evaluation and compared to the testing set which corresponded to the last 14 days of the observed data. Due to model weights being randomly set when defining the model and the stochastic nature of gradient descent these model outputs were not the same each time the program was run. Therefore, the program was run 50 times and the accuracy of these outputs evaluated both step-wise and borough-wise over each re-run using the R-squared (R^2), Root Mean Squared Error (RMSE) and Normalized Mean Squared Error (NRMSE) metrics to yield mean step-wise and borough-wise metrics representative across the 50 re-runs.

The R^2 metric measures the goodness of fit between two variables by measuring the proportion of the variance in the dependent variable that is predictable from the independent variable. R^2 is the square of r which is calculated as follows:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \quad \#(14)$$

Where:

- r is the correlation coefficient
- n is the number in the given dataset
- x is the first variable
- y is the second variable

The RMSE metric is similar to MSE in that it measures distances between a target variable and predicted values, only is equivalent to the square root of MSE:

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (Y_t - \bar{Y}_t)^2}{n}} \quad \#(15)$$

Where:

- Y_t is the actual value of a point for a given time period t
- n is the total number of fitted points
- \bar{Y}_t is the fitted forecast value for t

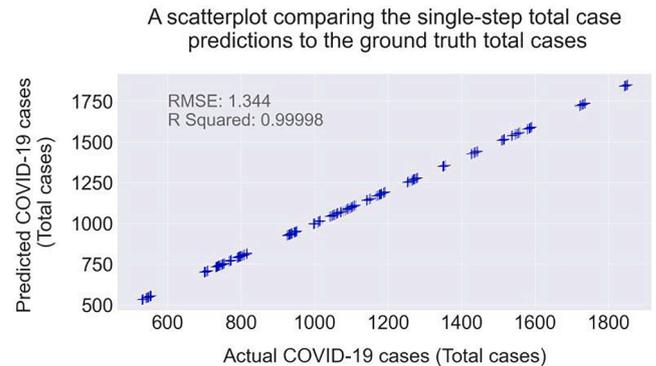


Figure 4. A scatterplot showing all of the single-step predicted values for all boroughs against their actual values

The third metric is NRMSE, which is equivalent to RMSE, but with the values normalized by dividing by the range of observed y values. This aids comparison between boroughs because it represents error proportional to magnitude of change:

$$NRMSE = \frac{RMSE}{R_y} \quad \#(16)$$

Where:

- R_y is the range of observed y values

The novel MSE-Moran's I loss function was also investigated further with a step-wise evaluation which compared the mean step-wise RMSE of the model when it was trained with the θ value at 0.003 (base model) versus 0 (thus effectively disabling the Moran's I component and reverting the loss function back to that of standard MSE), again calculated across the 50 re-runs.

Furthermore, a sensitivity analysis was conducted which systematically ablated various features and components from the model (except for the COVID-19 data which could not be removed due to the structure of the model) and evaluated the model across 50 re-runs in their respective absence. This analysis was conducted in order to better understand the relative importance of the variables used with regards to their contribution towards the spread of the disease.

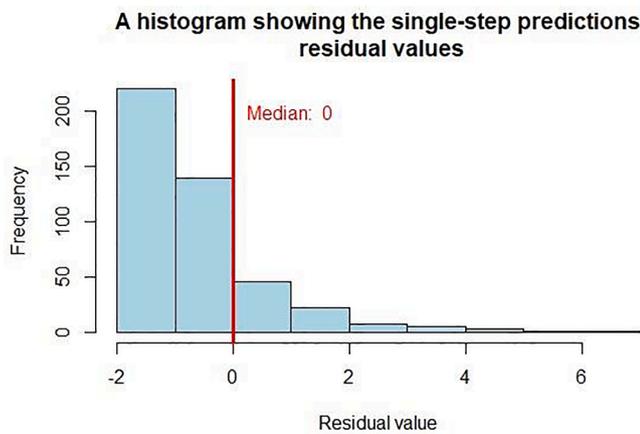


Figure 5. A histogram showing the distribution of the residuals calculated for all singlestep borough-level predictions for all time steps

Results

Single-step forecasting results

Overall, the single-step forecasts appear to be fairly accurate. This is evidenced by the high R_2 value for the whole of London (sum of all borough forecasts at each time step) total case forecasts (Figure 3 as well as the visually clear similarity between the true and predicted values.

Visualisation of every single-step prediction (448 total) for all individual boroughs in Figure 2 also communicates a similar message of accuracy to Figure 4 with an R_2 of 0.999 (3 d.p) and an RMSE of 1.344 (3 d.p) between the observed and predicted total case values.

Analysis of the residuals (Figure 5 and Table 1 of all 14 single-step forecasts made for each borough (448 in total) also suggest considerable accuracy with a median value of 0 and an interquartile range (IQR) of 1 stretching between -1 to 0, indicating that half of the predictions are either +1 or equal to the actual values. However, despite the median value of 0, the residual analysis does indicate a tendency for the model to slightly overpredict single-step forecasts. This is evidenced by the IQR, stretching from -1 to 0, but also by the actual residual distribution, with 220 residual values being negative and just 88 positive (with 140 equalling 0). This may largely be due to the dramatic overprediction through steps 11-14 where the actual number of new cases declines from

31 towards 0 (refer back to Figure 3.

Summary of results

This research proposed a hybrid LSTM-MLP deep learning model which is capable of providing borough-level short and long term COVID-19 forecasts within the city of London. This model was applied to produce 14 single-step forecasts as well as 14-day multi-step forecasts for each borough, the evaluation metrics of which are summarised in Tables 2,3 and 4.

Model predictions

Step-wise analysis

Generally, the model output step-wise RMSE (Figure 6 increases across the time-steps indicating decreasing accuracy. The proportion by which the RMSE increases across each time-step also appears roughly constant. However, this is with the exception of steps 7-9, whereby the increase in RMSE between steps is far lower than observed between time-steps elsewhere.

Borough-wise analysis

Regarding the borough-wise 14 day forecast RMSE (Figure 7a and Table 5, the mean and median RMSE values recorded across all boroughs were 100.120 and 95.991 respectively. It is worth noting that visually there generally appears to be an East-West axis of increasing RMSE whereby Western boroughs often exhibit a higher RMSE than the Eastern counterparts. This also evident with the two worst performing boroughs being situated in the West, with Ealing and Richmond upon Thames exhibiting the highest RMSE of 240.915 and 176.645 respectively. Equally, the best performing borough was also situated in the East, with Bexley exhibiting an RMSE of 36.418, although the next best performing borough was situated in the South with Sutton exhibiting an RMSE of 49.321.

Regarding the error relative to the magnitude of change (Figure 7b, the mean and median NRMSE values recorded across the borough-wise 14-day forecasts were 0.177 and 0.164 respectively. Similar to the RMSE metric, an East-West axis can be observed visually with the NRMSE generally increasing towards the West. Consistent with the RMSE, the two worst performing boroughs are situated in the West with Richmond upon Thames and Kingston upon Thames exhibiting the highest NRMSE of 0.311 and 0.259 respectively. The two best performing boroughs are also in the Eastern direction, with Waltham Forest (North-East) recording the lowest NRMSE of 0.106 and Bexley (East) recording the

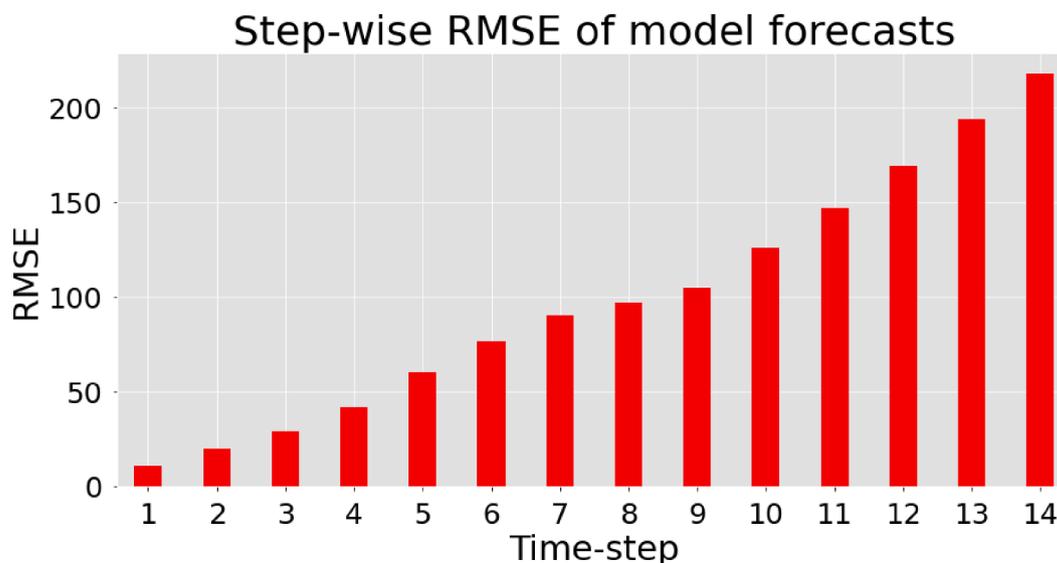


Figure 6. Showing the change in RMSE observed across each time-step.

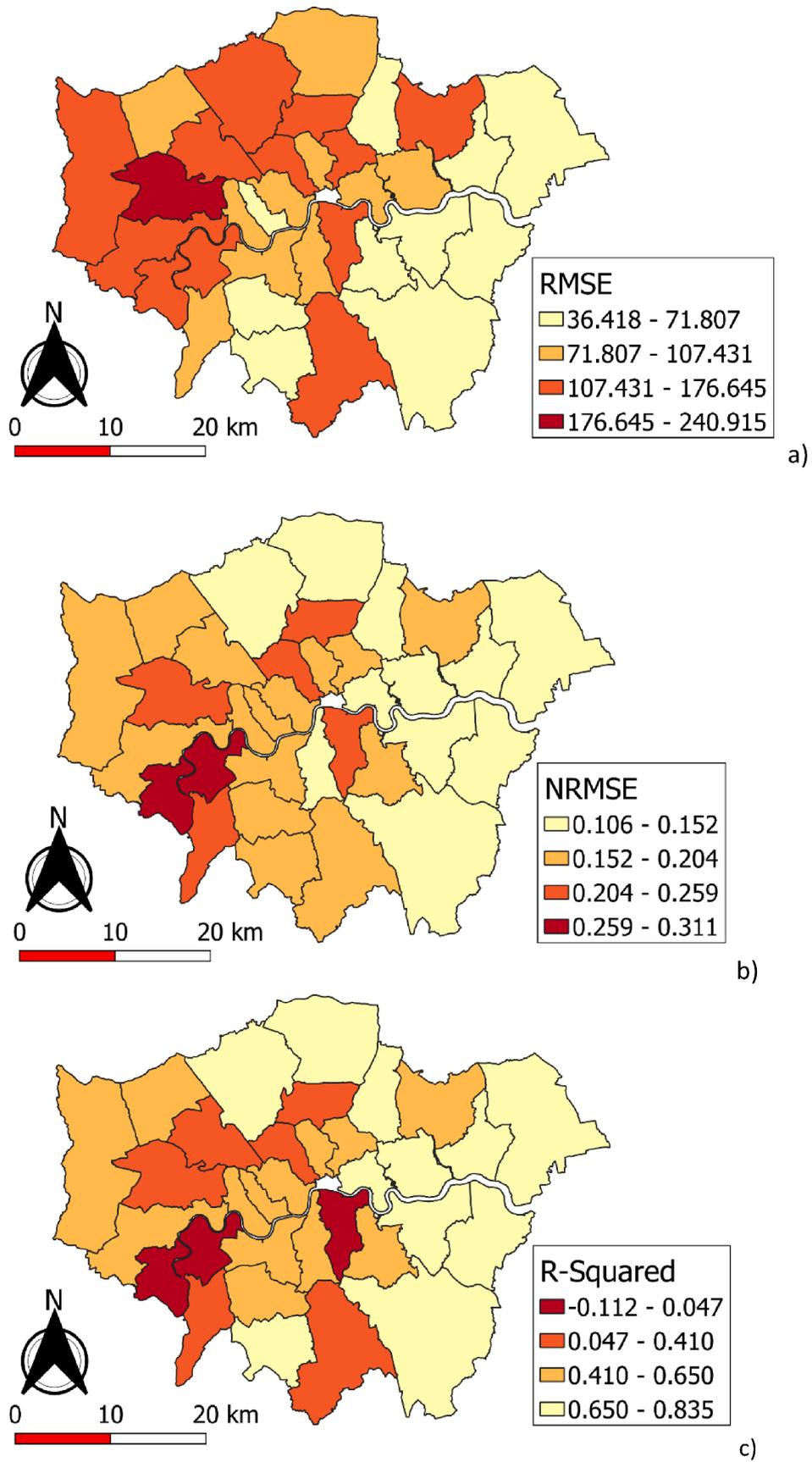


Figure 7. Maps showing the mean borough-wise RMSE (a), NRMSE (b) and R^2 (c) metrics.

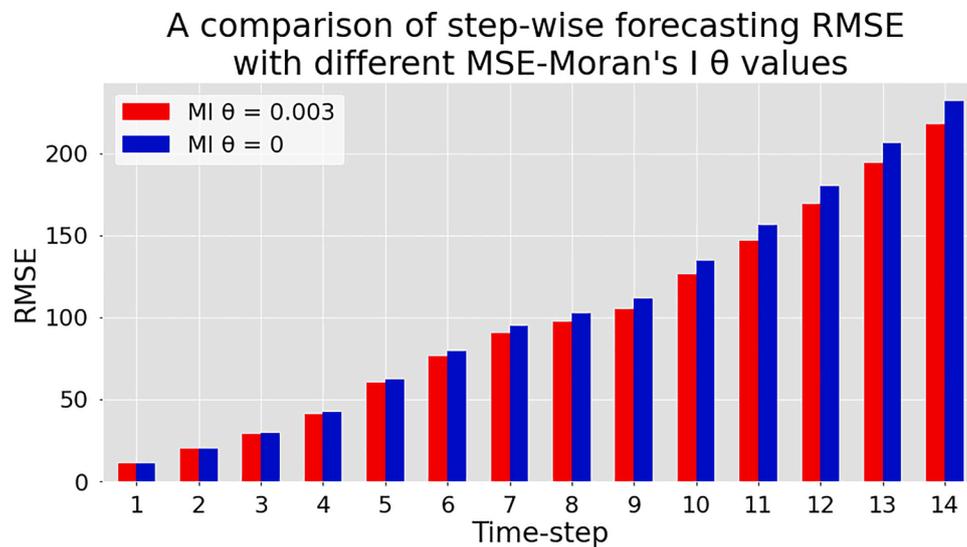


Figure 8. Showing a step-wise comparison of the model outputs with different MSE-Moran’s I loss function θ values.

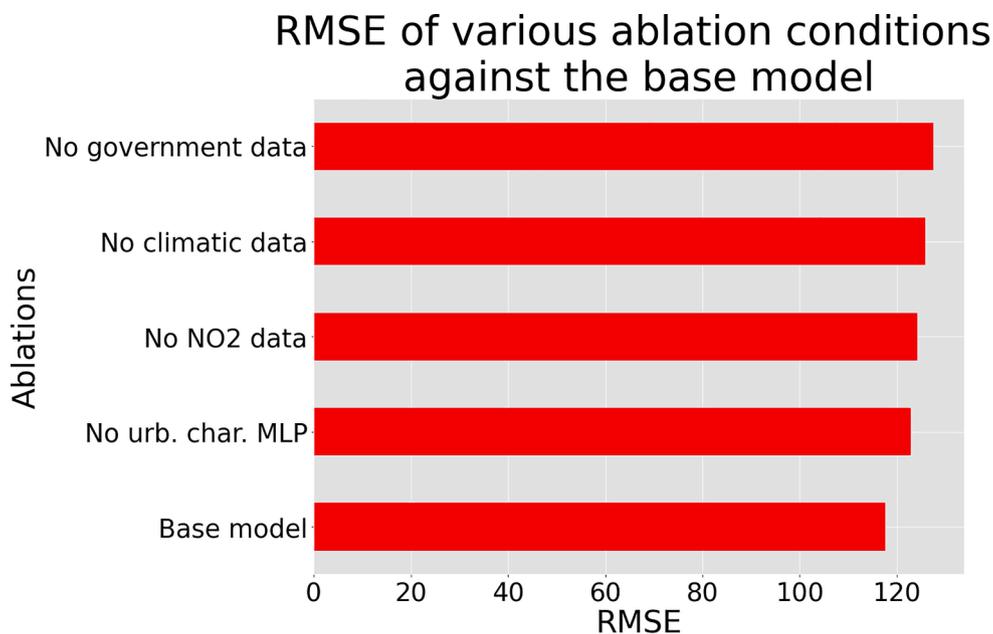


Figure 9. Comparing the mean output RMSE from 50 re-runs with the ablation of respective parts of the model against the base model.

next lowest at 0.112.

With respect to the goodness of fit (Figure 7c, the mean and median R^2 values were 0.547 and 0.609 respectively. Consistent with both the RMSE and NRMSE, an East-West axis is visually clear with Western boroughs exhibiting lower R^2 values indicating a poorer goodness of fit between the observed and forecasted values. Inspection of the worst performing boroughs also broadly follows this axis, with the worst performing borough being Richmond upon Thames (West) with an R^2 of -0.112, although the next worst performing borough is situated centrally with Southwark recording an R^2 of 0.047. The best performing boroughs also broadly follow this trend, with the best performing borough being Waltham Forest (North-East), and the next best being Bexley (East) with R^2 values of 0.835 and 0.822 respectively.

MSE-Moran's I loss function

The use of the MSE-Moran’s I loss function with θ set to 0.003 overall appears to improve the accuracy of forecasts (Figure 8 when compared

to θ at 0 (thus effectively removing the Moran’s I component and reverting the loss function back to that of standard MSE). Specifically, improvements range from 0.596% (step 1) to 6.274% (step 10), with improvements being observed across all time-steps with the exception of step 2 whereby the addition of the Moran’s I component to the loss function actually resulted in a small 1.528% increase in RMSE compared to the outputs with θ at 0.

Sensitivity analysis

Regarding the systematic ablation of various parts of the model (Figure 9, the removal of (i) the urban characteristics MLP, (ii) the NO_2 data, (iii) the climatic data, and (iv) the government stringency data all appear to result in lower output accuracy than the base model. Of all parts analysed, the removal of the government stringency index from the LSTM component appears to result in the greatest decrease in accuracy, with an 8.394% increase in RMSE. This is closely followed by the climatic data (7.031% increase in RMSE) and then NO_2 data (5.627%

increase in RMSE), and finally the urban characteristics component (4.483% increase in RMSE).

Hyperparameters

The hyperparameters which optimised the model's performance are summarised in Table 6.

Discussion

Model performance

As mentioned in the results, the step-wise RMSE of the model's outputs increases as the time-step increases. This is perhaps to be expected as model output inaccuracy accumulates and uncertainty increases as the time-step increases. However, despite the increase between steps generally appearing proportional, it is interesting to note that the increase in RMSE between steps 7 and 9 is considerably less than the others. Regarding the borough-level forecasts, since the model is the first to provide borough-level forecasts up to 14 days into the future, it remains somewhat difficult to definitively state how accurate these forecasts are since no comparable borough-level predictive modelling studies exist. However, the mean and median R^2 values of 0.547 and 0.609 recorded amongst boroughs as well as the high accuracy amongst many boroughs despite being expected to predict highly noisy data for up to 14 time-steps into the future may suggest that this model has considerable predictive potential. It is also interesting to note that there is a clear East-West axis amongst borough forecasting accuracy, whereby the model appears to provide a more accurate 14 days forecast for the boroughs situated in the East of London than the west.

The decrease in RMSE observed over 13 of the 14 time-steps with the MSE-Moran's I loss function θ value at 0.003 (base model) versus 0 (disabling the Moran's I component) also justifies the inclusion of the novel loss function as it indicates improved accuracy. It is interesting to note that the RMSE recorded was higher with the Moran's I component enabled at step 2, indicating reduced accuracy. However, given that this increase in RMSE was small (1.528%) and only took place at one time-step it is possible that this could be due to noise.

With respect to the sensitivity analysis, the increase in RMSE incurred by the removal of various components compared to the base model justifies their inclusion. Furthermore, the ordering of the increase in error incurred when parts were removed by ablation may also reflect the importance of features, with the government stringency data being the most important (excluding the COVID-19 data), followed by the climatic data, the NO_2 data and finally the urban characteristics component, which represented the least important part of the model.

Model implications

This research and the development of this model has implications. The first, and perhaps most obvious is that it fills a research gap by being the first explicitly designed to provide COVID-19 forecasts at a spatial scale lower than the city, such as at borough/district level and thus contributes to existing COVID-19 forecasting literature. Second, in a more applied context, this research and model may have the potential to aid governing bodies within cities in responding to COVID-19 by informing decision-making and enabling the vulnerable areas of a city to receive support at the right time.

This research also has implications at the intersection between deep learning and disease modelling because it introduced a novel MSE-Moran's I loss function which was overall demonstrated to improve the model's predictive potential. Furthermore, it may also have implications at the intersection of deep learning and spatio-temporal modelling more generally because it may have the potential to also improve the predictive potential of other spatio-temporal deep learning models if spatial autocorrelation is present. However, this remains to be proven and so

could warrant future investigation.

Model limitations

As with many ML experiments, the performance of this framework is affected by the quality of the training data available which represents a limitation. With this in mind, it should be noted that the borough-level COVID-19 data represents the number of COVID-19 cases confirmed by Pillar 1 and Pillar 2 testing. Therefore, whilst the data values are proportional to the prevalence of COVID-19 over time, it is also proportional to the availability of tests/ the number of tests being conducted over time. This is problematic because the testing capacity has often not met demand (Wise, 2020; Iacobucci, 2020), meaning that particularly in the early periods of COVID-19 waves the data may be under-representative. Furthermore, the number of tests being conducted has increased overtime (Gov, 2020), with considerably fewer tests being conducted during the first wave (roughly prior to June) compared to the second (roughly September/October), which makes modelling the dynamics of the second wave increasingly difficult using data from the first wave. This is especially problematic for the evaluation of this model because the testing set fell within the early stages of the second wave.

Future work

As mentioned, this model made use of a novel MSE-Moran's I loss function which improved the model's forecasts over all time-steps except 2 (although this may be noise). With this in mind, a future research opportunity could be to further experiment with this loss function/concept with other spatio-temporal datasets to both further validate it as an effective feature of spatio-temporal forecasts, but also to experiment with different variations/methods of deploying such a component. Also, regarding the performance of this model, whilst the metrics may suggest that this model has good predictive potential, the absence of sub-city level studies means that the predictive potential cannot be determined conclusively. Therefore, future research opportunities could also involve attempting to provide sub-city level forecasts but using different models and comparing the two. Another future avenue of work could also be to apply a similar model to this one within other cities perhaps of different sizes or in different parts of the world which have been affected by COVID-19 in order to further evaluate this framework.

Conclusions

This article introduced a novel LSTM-ANN deep learning model capable of producing borough-level COVID-19 forecasts up to 14 days into the future for the city of London, thus fundamentally addressing a research gap within COVID-19 modelling literature by being the first model designed and optimized explicitly to forecast future values at borough-level within a city. Specifically, the model was designed to do so by considering temporally lagged borough-level COVID-19 data, as well as temporally lagged borough-level NO_2 concentrations, government stringency data, and climatic data, and additionally the urban characteristics of each borough. The model was also encouraged to learn the spatial relationships between boroughs with regards to the spread of COVID-19 by a novel MSE-Moran's I loss function applied in the training process.

Owing to the lack of comparable studies due to this model being the first of its kind, it remains difficult to confidently state just how accurate it is. However, the step-wise and borough-wise model evaluation metrics show promise despite the model being expected to predict noisy data for up to 14 time-steps into the future, suggesting that this model has considerable predictive potential. As such, the model represents a useful tool for assisting the decision making and interventions of governing bodies within cities which remains the main implication of this research. Also, the sensitivity analysis indicates that the government stringency

data is particularly important in the modelling process, with this being closely followed by the climatic variables, the NO₂ concentration data, and finally the urban characteristics data. Furthermore, this research also introduced a novel MSE-Moran's I loss function which was demonstrated to improve the forecasting accuracy across all 14 future time-steps with the exception of the second. As such, this research has implications at the intersection of deep learning and disease epidemiological modelling. Such a function may also have secondary implications at the intersection of deep learning and spatio-temporal analytics more generally because this technique may help improve the accuracy in other spatio-temporal forecasting applications. [1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70]

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Abbasimehr H, Shabani M, Yousefi M. An optimized model using LSTM network for demand forecasting. *Computers & Industrial Engineering* 2020;106435.
- Achuo ED. How efficient are government stringency responses in curbing the spread of the covid-19 pandemic? *International Journal of Research and Innovation in Social Science* 2020;4(8).
- Ahmadi, M., Sharifi, A., Dorosti, S., Ghouschi, S.J. and Ghanbari, N., 2020. Investigation of effective climatology parameters on COVID-19 outbreak in Iran. *Science of the Total Environment*, p.138705.
- Balli S. Data analysis of Covid-19 pandemic and short-term cumulative case forecasting using machine learning time series methods. *Chaos, Solitons & Fractals* 2020;142:110512.
- Carteni A, Di Francesco L, Martino M. How mobility habits influenced the spread of the COVID-19 pandemic: Results from the Italian case study. *Science of the Total Environment* 2020;741:140489.
- Che Z, Purushotham S, Cho K, Sontag D, Liu Y. Recurrent neural networks for multivariate time series with missing values. *Scientific reports* 2018;8(1):1–12.
- Chimmula VKR, Zhang L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos, Solitons & Fractals* 2020:109864.
- Dehesh, T., Mardani-Fard, H.A. and Dehesh, P., 2020. Forecasting of covid-19 confirmed cases in different countries with arima models. *medRxiv*.
- De Castro, C.A., 2020. SIR Model for COVID-19 calibrated with existing data and projected for Colombia. *arXiv preprint arXiv:2003.11230*.
- DT, 2020. Licensed Vehicles – Type, Borough. Published online at: [Licensed Vehicles - Type, Borough - London Datatore](https://data.london.gov.uk/dataset/licensed-vehicles-type-borough-london-datatore). [Online; accessed 23rd October 2020].
- Erraguntla M, Zapletal J, Lawley M. Framework for Infectious Disease Analysis: A comprehensive and integrative multi-modeling approach to disease prediction and management. *Health informatics journal* 2019;25(4):1170–87.
- Fanelli D, Piazza F. Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos, Solitons & Fractals* 2020;134:109761.
- Feng, C. and Jiao, J., Predicting and mapping neighborhood-scale health outcomes: A machine learning approach. *Computers, Environment and Urban Systems*, 85, p.101562.
- GLA, 2020a. Coronavirus (COVID-19) Cases. Published online at: <https://data.london.gov.uk/dataset/coronavirus-covid-19-cases>. [Online; accessed 23rd October 2020].
- GLA, 2020b. London Borough Profiles and Atlas. Published online at: <https://data.london.gov.uk/dataset/london-borough-profiles>. [Online; accessed 23rd October 2020].
- GLA, 2020c. Statistical GIS Boundary Files for London. Published online at: <https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london>. [Online; accessed 23rd October 2020].
- Goscé L, Phillips A, Spinola P, Gupta RK, Abubakar I. Modelling SARS-COV2 Spread in London: Approaches to Lift the Lockdown. *Journal of Infection*. 2020.
- Gov, 2020. Coronavirus (COVID-19) in the UK. Published online at: <https://coronavirus.data.gov.uk/details/testing>. [Online; accessed 10th December 2020].
- Greff K, Srivastava RK, Koutnik J, Steunebrink BR, Schmidhuber J. LSTM: A search space odyssey. *IEEE transactions on neural networks and learning systems* 2016;28(10):2222–32.
- Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical characteristics of coronavirus disease 2019 in China. *New England journal of medicine* 2020;382(18):1708–20.
- Hale T, Webster S, Petherick A, Phillips T, Kira B. Oxford covid-19 government response tracker. *Blavatnik School of Government* 2020;25.
- Han J, Pei J, Kamber M. Data mining: concepts and techniques. Elsevier; 2011.
- Haykin S. *Neural Networks and Learning Machines*. Pearson Education India; 20103/E..
- Ho, F.K., Celis-Morales, C.A., Gray, S.R., Katikireddi, S.V., Niedzwiedz, C.L., Hastie, C., Lyall, D.M., Ferguson, L.D., Berry, C., Mackay, D.F. and Gill, J.M., 2020. Modifiable and non-modifiable risk factors for COVID-19: results from UK Biobank. *medRxiv*.
- Hyndman RJ, Athanasopoulos G. Chapter 8 ARIMA models. *Forecasting: principles and practice*. Melbourne, Australia: OTexts; 2018.
- Iacobucci G. Covid-19: Testing service wasn't prepared for increased demand, chief admits. *BMJ: British Medical Journal (Online)* 2020;370.
- Ibrahim, M.R., Haworth, J., Lipani, A., Aslam, N., Cheng, T. and Christie, N. 2020. Variational-LSTM Autoencoder to forecast the spread of coronavirus across the globe. *medRxiv*.
- Johnson T, McQuarrie B. Mathematical modeling of diseases: Susceptible-infected-recovered (sir) model. Morris, Math 4901 Senior Seminar: In University of Minnesota; 2009.
- Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character* 1927;115(772):700–21.
- Kim S, Castro MC. Spatiotemporal pattern of COVID-19 and government response in South Korea (as of May 31, 2020). *International Journal of Infectious Diseases* 2020;98:328–33.
- Lake MA. What we know so far: COVID-19 current clinical knowledge and research. *Clinical Medicine* 2020;20(2):124.
- LAQN, 2020. London Air data downloads. Published online at: <http://www.londonair.org.uk/london/asp/dataspecies.asp?site1=&site2=&site3=&site4=&site5=&site6=&day1=1&month1=jan&year1=2020&day2=18&month2=jun&year2=2020&period=daily&species=NO2&ratidate=&res=6&Submit=Replot+graph>. [Online; 23rd October 2020].
- Lee N. Inclusive Growth in cities: a sympathetic critique. *Regional Studies* 2019;53(3):424–34.
- Mao L, Bian L. Spatial-temporal transmission of influenza and its health risks in an urbanized area. *Computers, Environment and Urban Systems* 2010;34(3):204–15.
- Mei S, Chen B, Zhu Y, Lees MH, Boukhanovsky AV, Sloot PM. Simulating city-level airborne infectious diseases. *Computers, Environment and Urban Systems* 2015;51:97–105.
- Melo GCD, Araújo Neto RAD, Araújo KCGMD. Forecasting the rate of cumulative cases of COVID-19 infection in Northeast Brazil: a Boltzmann function-based modeling study. *Cadernos de Saúde Pública* 2020;36:e00105720.
- MHCLG, 2020. Total Number of Dwellings and Net Additional Dwellings, Borough. Published online at: <https://data.london.gov.uk/dataset/net-additional-dwellings-borough>. [Online; 23rd October 2020].
- NOAA (2020) Climate Data Online Search. Available at: <https://www.ncdc.noaa.gov/cdo-web/search>. [Online; 23rd October 2020].
- Parbat D, Chakraborty M. A python based support vector regression model for prediction of COVID19 cases in India. *Chaos, Solitons & Fractals* 2020;138:109942.
- Parmezan ARS, Souza VM, Batista GE. Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model. *Information Sciences* 2019;484:302–37.
- Petropoulos F, Makridakis S. Forecasting the novel coronavirus COVID-19. *PLoS one* 2020;15(3):e0231236.
- Pope, A., 2017. Location of Supermarkets in Britain. Published Online at: <https://datastore.is.ed.ac.uk/handle/10283/2413?show=full>. [Online; 23rd October 2020].
- Prats-Urbe, A., Paredes, R. and Prieto-Alhambra, D., 2020. Ethnicity, comorbidity, socioeconomic status, and their associations with COVID-19 infection in England: a cohort analysis of UK Biobank data. *medRxiv*.
- Ramasamy D, Chinnu G. COVID-19 Situation at Chennai City-Forecasting for the Better Pandemic Management. *Journal of Current Research and Review* 2020;12(12):37–47.
- Ribeiro MHD, da Silva RG, Mariani VC, dos Santos Coelho L. Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil. *Chaos, Solitons & Fractals* 2020:109853.
- Ritchie, H., 2018. Urbanization. Published online at <http://ourworldindata.org/>. Retrieved from <https://ourworldindata.org/urbanization>. [Online; accessed 3rd June 2020].
- Roosa K, Lee Y, Luo R, Kirpich A, Rothenberg R, Hyman JM, et al. Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. *Infectious Disease Modelling* 2020;5:256–63.
- Scarpino SV, Petri G. On the predictability of infectious disease outbreaks. *Nature communications* 2019;10(1):1–8.
- Semenova Y, Glushkova N, Pivina L, Khismetova Z, Zhunussov Y, Sandybaev M, et al. Epidemiological characteristics and forecast of COVID-19 outbreak in the Republic of Kazakhstan. *Journal of Korean medical science* 2020;35(24).
- Shastri S, Singh K, Kumar S, Kour P, Mansotra V. Time series forecasting of Covid-19 using deep learning models: India-USA comparative case study. *Chaos, Solitons & Fractals* 2020;140:110227.
- Sugishita Y, Kurita J, Sugawara T, Ohkusa Y. Forecast of the COVID-19 outbreak, collapse of medical facilities, and lockdown effects in Tokyo. *Japan. medRxiv*. 2020.
- Thara DK, PremaSudha BG, Xiong F. Auto-detection of epileptic seizure events using deep neural network with different feature scaling techniques. *Pattern Recognition Letters* 2019;128:544–50.
- Toda, A.A., 2020. Susceptible-infected-recovered (sir) dynamics of covid-19 and economic impact. *arXiv preprint arXiv:2003.11221*.

- [54] Ture M, Kurt I. Comparison of four different time series methods to forecast hepatitis A virus infection. *Expert Systems with Applications* 2006;31(1):41–6.
- [55] Vapnik V. *The nature of statistical learning theory*. Springer science & business media; 2013.
- [56] Voyant C, Notton G, Kalogirou S, Nivet ML, Paoli C, Motte F, et al. Machine learning methods for solar radiation forecasting: A review. *Renewable Energy* 2017;105:569–82.
- [57] Wang P, Zheng X, Ai G, Liu D, Zhu B. Time series prediction for the epidemic trends of COVID-19 using the improved LSTM deep learning method: Case studies in Russia, Peru and Iran. *Chaos, Solitons & Fractals* 2020;140:110214.
- [58] Wesolowski M, Suchacz B. Artificial neural networks: theoretical background and pharmaceutical applications: a review. *Journal of AOAC International* 2012;95(3): 652–68.
- [59] Wise J. Covid-19: What's going wrong with testing in the UK?. 2020;bmj:370.
- [60] WHO, 2020. Virtual press conference on COVID-19 – 11 March 2020. https://www.who.int/docs/default-source/coronaviruse/transcripts/who-audio-emergencies-coronavirus-press-conference-full-and-final-11mar2020.pdf?sfvrsn=cb432bb3_2. [Online; accessed 3rd June 2020].
- [61] Worldometer, 2020. COVID-19 Coronavirus Pandemic. <https://www.worldometers.info/coronavirus/> [Online; accessed 22nd December 2020].
- [62] Wu CH, Ho JM, Lee DT. Travel-time prediction with support vector regression. *IEEE transactions on intelligent transportation systems* 2004;5(4):276–81.
- [63] Yan, B., Tang, X., Liu, B., Wang, J., Zhou, Y., Zheng, G., Zou, Q., Lu, Y. and Tu, W., 2020. An Improved Method of COVID-19 Case Fitting and Prediction Based on LSTM. *arXiv preprint arXiv:2005.03446*.
- [64] Yang Y, Atkinson P, Ettema D. Individual space–time activity-based modelling of infectious disease transmission within a city. *Journal of the Royal Society Interface* 2008;5(24):759–72.
- [65] Yang Z, Zeng Z, Wang K, Wong SS, Liang W, Zanin M, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of Thoracic Disease* 2020;12(3):165.
- [66] Yonar H, Yonar A, Tekindal MA, Tekindal M. Modeling and Forecasting for the number of cases of the COVID-19 pandemic with the Curve Estimation Models, the Box-Jenkins and Exponential Smoothing Methods. *EJMO* 2020;4(2):160–5.
- [67] Kizilcec V, et al. Forecasting Solar Home System Customers' Electricity Usage with a 3D Convolutional Neural Network to Improve Energy Access. *Energies* 2022. <https://doi.org/10.3390/en15030857>.
- [68] Shelbourne C, et al. Inference of virtual network functions' state via analysis of the CPU behavior. 33th International Teletraffic Congress (ITC-33) 2022. <http://ieeexplore.ieee.org/abstract/document/9625619>.
- [69] Anderson-Bell J, et al. Predicting non-residential building fire risk using geospatial information and convolutional neural networks. *Remote Sensing Applications: Society and Environment* 2021. <https://doi.org/10.1016/j.rsase.2021.100470>.
- [70] James T, et al. Convolutional neural networks for water segmentation using sentinel-2 red, green, blue (RGB) composites and derived spectral indices. *International Journal of Remote Sensing* 2021. <https://doi.org/10.1080/01431161.2021.1913298>.