Integrating LSTMs and GNNs for COVID-19 Forecasting

Nathan Sesti [* 1] **Juan Jose Garau-Luis** [* 1] **Edward Crawley** [1] **Bruce Cameron** [1]

Abstract

The spread of COVID-19 has coincided with the rise of Graph Neural Networks (GNNs), leading to several studies proposing their use to better forecast the evolution of the pandemic. Many such models also include Long Short Term Memory (LSTM) networks, a common tool for time series forecasting. In this work, we further investigate the integration of these two methods by implementing GNNs within the gates of an LSTM and exploiting spatial information. In addition, we introduce a skip connection which proves critical to jointly capture the spatial and temporal patterns in the data. We validate our daily COVID-19 new cases forecast model on data of 37 European nations for the last 472 days and show superior performance compared to state-of-the-art graph time series models based on mean absolute scaled error (MASE). This area of research has important applications to policy-making and we analyze its potential for pandemic resource control.

1. Introduction

Since its outbreak in late 2019, the COVID-19 virus (Fauci et al., 2020; Velavan & Meyer, 2020) has devastated the world, causing over 3 million deaths (World Health Organization, 2021) and economic losses estimated up to $10 trillion (United Nations, 2020). Crucial to curbing such damage is prompt response and effective policy-making, and, even with vaccination numbers increasing, this ability will be crucial to develop as the possibility of a future pandemic looms. Due to the exponential nature of epidemiological transmission (Li et al., 2020), even slight improvements on early intervention can have outsized impact, making preemptiveness one of our best tools. Therefore, knowledge of future spread is of critical importance.

Given the urgency of the pandemic, the Machine Learning (ML) community has stepped in to help in multiple capacities (Alimadadi et al., 2020; Buckee et al., 2020), from assisting in early COVID-19 diagnosis based on CT scans (Barstugan et al., 2020; Wang et al., 2021), to predicting new cases and hospitalizations using different time series models (Alazab et al., 2020). Among the latter efforts, some authors have taken advantage of both the temporal correlations and the inherent network structure of epidemiological data and thus combined temporal models such as Long Short Term Memory (LSTM) networks (Hochreiter & Schmidhuber, 1997) with recent advancements in Graph Neural Networks (GNNs) (Scarselli et al., 2008). This has proven to be one of the most successful approaches.

In this work we aim to improve upon their accuracy, presenting a new time series forecasting model based on GNNs and LSTMs. We introduce a general method that further integrates both networks and apply it to the problem of forecasting new COVID-19 cases. Our specific use case consists of 37 countries in Europe, with data from Jan. 2020 to May 2021. We show that our model is able to forecast new cases out of the training distribution, showing better results than four other GNN+LSTM models and lag, with a prediction error of 10% when forecasting 7 days ahead. We also analyze how the model performance translates into real-world policy-making by means of the fraction of missed cases. Finally, our findings suggest that introducing a skip connection in our model is key to reducing underfitting.

2. Related Work

Following the outbreak of COVID-19, the ML community has given the problem of epidemiological modeling a refreshed solution. The majority of the studies have made LSTMs the tool of choice. In addition, many researchers have answered the call for use of mobility data (Buckee et al., 2020; Santosh, 2020) and used graphs to model spatial information, some converging upon different combinations of LSTMs and GNNs to solve the problem, mirroring their use in other applications.

Temporal-Only Models Since early in the pandemic, researchers have realized the value of LSTM networks due to the salient temporal patterns of transmission (Chimmula & Zhang, 2020; Shahid et al., 2020; Arora et al., 2020). Their predictions displayed low error, competing well with stan-
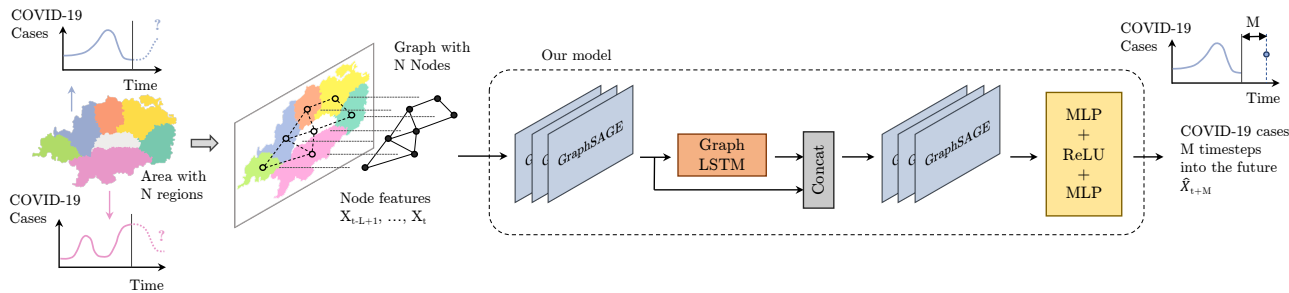
---

*Equal contribution [1]Engineering Systems Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA. Correspondence to: Juan Jose Garau-Luis <garau@mit.edu>.

Figure 1. Overview of the problem and our proposed method. We consider an area with N different regions (e.g., different countries, states), each displaying a different COVID-19 case curve. The task consists of predicting, M timesteps in advance (we use days in our paper), the number of cases in each region. To that end, a graph with N nodes is created based on the geographical structure of the regions and a feature vector is associated to each node. These are fed to our model, consisting of three layers of GraphSAGE convolutions, an LSTM with graph convolution gates, three additional GraphSAGE layers, and two final fully-connected layers. We use a skip connection between the first batch of GraphSAGE layers and the GraphLSTM.

dard epidemiological models, such as susceptible-infected-recovered-dead (SIRD) models (Bailey et al., 1975).

GNNs for COVID-19 Capitalizing on the call for use of mobility data, numerous papers have applied GNNs to the task. Some have leveraged classic epidemiological models, one feeding the concatenation of outputs from a GNN and an epidemiological model into a final distributional regression layer (La Gatta et al., 2021). Alternatively, (Fritz et al., 2021) uses a combination of a GNN and LSTM to predict the parameters of a SIRD model. Our work will differ in that it is not dependent upon any explicit epidemiological model, instead being able to construct its own parametrization of input features.

While many studies take advantage of both the spatial and temporal dependencies of the problem, (Cao et al., 2021) pays particular attention to modeling them jointly — i.e., leveraging the correlations that depend on both other regions and timesteps rather than taking these separately. While they achieve this using Fourier Transforms, we instead draw from (Seo et al., 2018), which replaces the typical linear transformation within an LSTM with a Chebyshev Spectral CNN (Defferrard et al., 2016), although we switch it out with GraphSAGE (Hamilton et al., 2017). Our contribution also includes a novel skip connection, which concatenates the output of the aforementioned LSTM module with a purely spatial GNN in order to improve upon the challenges common to joint modeling of spatial and temporal dimensions.

Other Applications of GNN + LSTM At the intersection of these two network models has been an explosion of research aiming to tackle other graph-structured time series problems. Specifically, DCRNN (Li et al., 2017) was applied to traffic prediction, GCLSTM (Chen et al., 2018) was applied to predicting emails and other contact between people, and GConvGRU (Seo et al., 2018) was able to predict moving MNIST data and model natural language.

3. Problem Statement

Our objective in this paper is to predict some numerical feature of each node in a graph given a sequence of previous graph snapshots. In the case of COVID-19, we want to predict new cases across different regions, where each region constitutes a different node in the graph. The data can be characterized as a graph $\mathcal{G} = (X, A, W)$ where $X = \{x_{ijt}\} \in \mathbb{R}^{K_X \times N \times T}$ indicates the K_X input features for each of N nodes at T different timesteps, with $X_t \in \mathbb{R}^{K_X \times N}$ indicating the values at timestep t. Then, A is the adjacency matrix —fixed over time— and $W \in \mathbb{R}^{K_W \times N \times N}$ represents each of the K_W static edge features used in our model between each pair of countries.

The task at timestep t is, given the L most recent timesteps $\{X_{t-L+1}, \ldots, X_t\}$, to predict the number of new cases M timesteps into the future, $\hat{X}_{t+M} \in \mathbb{R}^N$. When each graph snapshot corresponds to a unique day, the task becomes predicting new cases M days ahead, which benefits effective policy-making and early response if M is large enough.

4. Methods

Our method consists of a combination of GraphSAGE and LSTM layers that jointly exploit spatio-temporal relationships in the data. This section covers the main details of the model, wich is depicted in Figure 1. Our approach considers edge features in the GraphSAGE layers and replaces linear operations in the LSTM by graph convolutions.

GraphSAGE Our model makes extensive use of Graph-SAGE (Hamilton et al., 2017), a spatial GNN that aggregates the features of local nodes to generate an embedding for each node in a graph, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Let $\boldsymbol{x}_j = \boldsymbol{h}_j^0$ represent the feature vector of node j, which is also the embedding for that node before any iterations. In each iteration $k \in \{1, ..., n\}$, each node $v \in \mathcal{V}$ aggregates the embeddings

of all immediately adjacent nodes $\{\boldsymbol{h}_u^{k-1}, \; \forall u \in \mathcal{N}(v)\}$, usually by taking the mean of these vectors. This aggregated vector, $\boldsymbol{h}_{\mathcal{N}(v)}^{k-1}$, is then concatenated to the current embedding $\boldsymbol{h}_v^{k-1}$ and fed into a single neural network layer with a sigmoid activation function to compute $\boldsymbol{h}_v^k$. We slightly alter this scheme to allow for edge features e_{vu} between any two nodes u and v, taking a weighted mean, $\boldsymbol{h}_{\mathcal{N}(v)}^{k-1} = \frac{1}{|\mathcal{N}(v)|} \sum_{u \in \mathcal{N}(v)} e_{vu} \cdot h_u^{k-1}$. In order to use multiple edge features, we learn to calculate the scalar e_{vu} from a feature vector $\boldsymbol{e}_{vu}$ using a single-layer perceptron.

GraphLSTM LSTMs (Hochreiter & Schmidhuber, 1997) are a class of recurrent neural architectures which use a sequence of gates to keep track of both short- and long-term dependencies in the data. Each of these gates usually applies weights via matrix multiplication; however, we instead replace each of these linear transformations with a graph convolution operation, specifically GraphSAGE, in order to maintain the spatial structure of our inputs and jointly model spatio-temporal dependencies. We refer to this modification of the LSTM architecture as GraphLSTM.

Skip Connection The overall structure of our model (see Figure 1) consists of an initial GraphSAGE layer with $n = 3$, feeding into a GraphLSTM with embedded GraphSAGE, followed by a final GraphSAGE layer also using $n = 3$. A multi-layer perceptron with ReLU activation is applied to finally generate a prediction. Rather than being fed only the output of the GraphLSTM, the final GraphSAGE layer recieves a concatenation of the output of the first GraphSAGE layer and the LSTM. This skip connection is an important contribution of our work, as it is found to speed stabilization of the model and discourage underfitting.

Refinements In addition to the presented structure, in both GraphSAGE layers, we make use of common ML techniques including: 1) dropout, which randomly disables nodes in the graph in order to discourage overfitting and increase model robustness (Srivastava et al., 2014), and 2) residual connections, which concatenate raw inputs to posterior layers in order to address the vanishing gradient problem (He et al., 2016).

Loss Function The specific loss function we use is mean absolute scaled error (MASE). To train the model we look at the per-person error, i.e., we aggregate all geographical regions and consider the absolute number of COVID-19 cases. Specifically,

$$\mathcal{L}_t = \frac{|\sum_{i=0}^{N-1} \hat{X}_{ti} - X_{ti}|}{\sum_{i=0}^{N-1} X_{ti}} \tag{1}$$

where N is the total number of geographical regions considered.

Table 1. Per-person and per-country MASE on the test data for our model and each of the 5 compared approaches, including Lag prediction.

MODEL	PER-PERSON	PER-COUNTRY
OUR MODEL	**0.10**	**0.27**
GCONVLSTM	0.83	2.15
GCONVGRU	0.78	1.59
DCRNN	0.83	2.15
GCLSTM	0.83	1.66
LAG	0.13	0.30

5. Results

We now apply our model[1] to the task of COVID-19 forecasting. We start by explaining the dataset and training procedure, and then discuss the results.

Dataset and Training Motivated by the dynamics of the virus in Europe (Dye et al., 2020; Saglietto et al., 2020), we select a use case consisting of daily new COVID-19 cases for 37 European nations ($N = 37$) with more than 100k inhabitants. We use new COVID-19 cases data from (Dong et al., 2020) for the period 24 Jan. 2020 - 9 May 2021. We use the first 377 days as training data, the following 47 days as validation data, and the last 48 days as test data. Our task consists of predicting the smoothed number[2] of new cases one week ahead ($M = 7$). We train our model by individually passing the data from each of 21 consecutive days ($L = 21$), beginning 28 days prior to our prediction target, and use the output of the 21st day as our prediction. The edges connecting countries are chosen based on proximity, each node being assigned to its 3 nearest neighbors, as determined by geodesic distance between landmass centroids. We use a single edge feature ($K_W = 1$), consisting of the social connectivity score between two incident nations according to (Facebook, 2021).

Baseline Comparison The results of running our model on the aforementioned data are shown in Figure 2. We compare our approach with four state-of-the-art models which have been shown effective in similar spatial time series forecasting problems: GConvLSTM (Seo et al., 2018), GConvGRU (Seo et al., 2018), DCRNN (Li et al., 2017), and GCLSTM (Chen et al., 2018). We find that ours displays significant improvements for the task of COVID-19 prediction. We also compare against lag, which uses the last input date as its prediction. This mechanism has strong results, and though our model learns a similar lagged behaviour, it yields a 20% reduction in loss over pure lag. Specific numbers are presented in Table 1 for both our loss function,

---

[1]A repository with our code can be found at: https://github.com/jjgarau/GNND
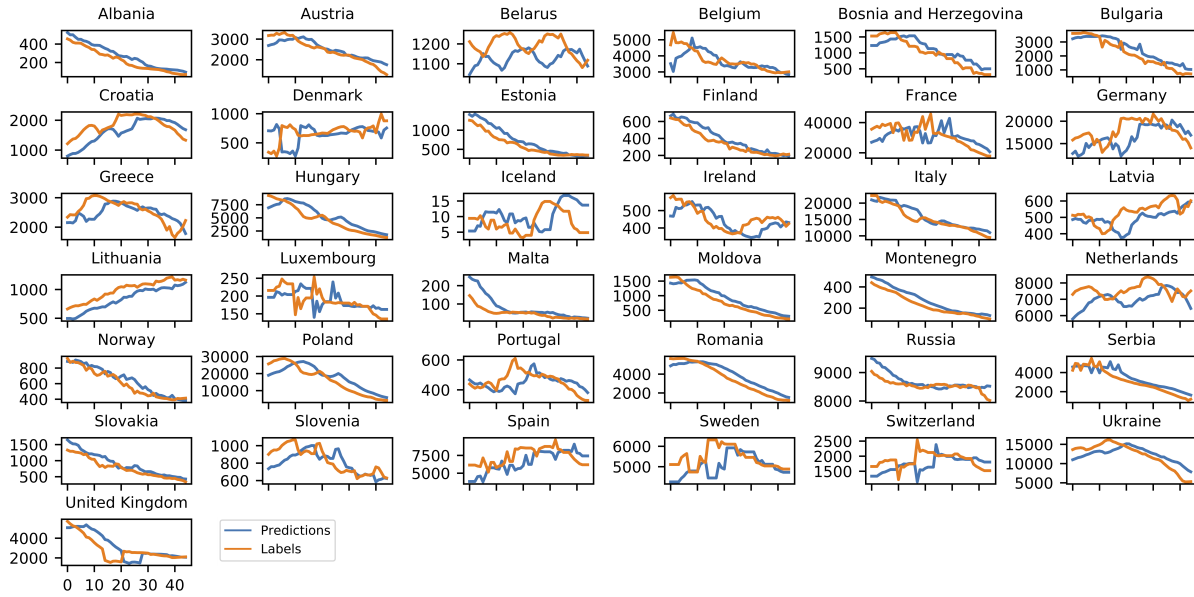
[2]Moving average over one week

Figure 2. Predictions and labels for the test data, consisting of new COVID-19 cases in 37 European countries for the period 22 Mar. 2021 - 9 May 2021 (48 days).

per-person MASE, and per-country MASE, which is instead calculated such that each nation weights equally regardless of its total number of cases.

Application to Policy-making Both of the aforementioned metrics are useful for policy-making: per-person MASE in cases of international policy-making, such as that done by the EU; and per-country MASE in the case where countries must put national policies in place. To better understand the real impact of our predictions, we also present an illustration of the fraction of missed cases (i.e., undercounting) in each country (Figure 3) when using our model on the 48 days of test data. Although time series models are usually trained with symmetric loss functions such as MASE or MSE, their real-world impact might be asymmetric, as in the case of COVID-19, where underprediction might hamper preparedness. In our case, the fraction of missed cases ranges from 0 to 0.16, and stays below 0.1 for the majority of the countries.

Skip Connection Ablation Study To validate the efficacy of our proposed skip connection, we perform an ablation test, finding that the altered model achieves an inferior result of 0.76 for per-person MASE (1.68 per-country MASE). We conclude that the skip connection plays a critical role in learning the spatio-temporal patterns present in the data.

6. Conclusion and Future Work

In this paper, we presented a model which built upon existing approaches to COVID-19 prediction by further inte-
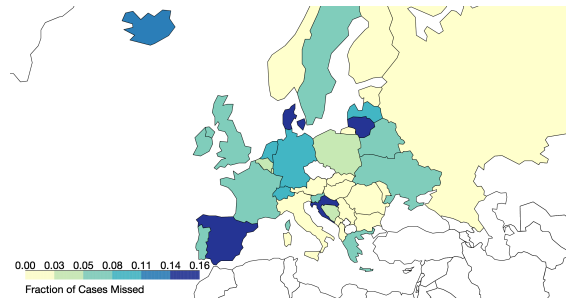

Figure 3. Fraction of missed cases on the 48 days of test data.

grating LSTMs and GNNs. The outputs of our model could prove useful to policy-makers attempting to take preemptive action by giving them improved knowledge of future evolution of the pandemic. Our model embeds the GraphSAGE graph convolution operator in place of linear transformations within the gates of an LSTM to create GraphLSTM, a module capable of jointly capturing spatio-temporal patterns. We also propose a skip connection to solve common challenges of time series modeling, which is shown to be an important addition to the model. Our work also presents a solution useful beyond the application of epidemiological prediction. Numerous problems similarly rely on a graph structure with node features changing in time. Future work may also further capitalize on choice of data—while we used only single edge and node features, there are certainly other relevant factors potentially including hospital capacity, poverty rates, and age demographics.

References

Alazab, M., Awajan, A., Mesleh, A., Abraham, A., Jatana, V., and Alhyari, S. Covid-19 prediction and detection using deep learning. *International Journal of Computer Information Systems and Industrial Management Applications*, 12:168–181, 2020.

Alimadadi, A., Aryal, S., Manandhar, I., Munroe, P. B., Joe, B., and Cheng, X. Artificial intelligence and machine learning to fight covid-19. *Physiological Genomics*, 52 (4):200–202, 2020.

Arora, P., Kumar, H., and Panigrahi, B. K. Prediction and analysis of covid-19 positive cases using deep learning models: A descriptive case study of india. *Chaos, Solitons & Fractals*, 139:110017, 2020. ISSN 0960-0779.

Bailey, N. T. et al. *The mathematical theory of infectious diseases and its applications*. Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE., 1975.

Barstugan, M., Ozkaya, U., and Ozturk, S. Coronavirus (COVID-19) Classification using CT Images by Machine Learning Methods. *arXiv preprint arXiv:2003.09424*, mar 2020.

Buckee, C. O., Balsari, S., Chan, J., Crosas, M., Dominici, F., Gasser, U., Grad, Y. H., Grenfell, B., Halloran, M. E., Kraemer, M. U. G., Lipsitch, M., Metcalf, C. J. E., Meyers, L. A., Perkins, T. A., Santillana, M., Scarpino, S. V., Viboud, C., Wesolowski, A., and Schroeder, A. Aggregated mobility data could help fight covid-19. *Science*, 368(6487):145–146, 2020. ISSN 0036-8075.

Cao, D., Wang, Y., Duan, J., Zhang, C., Zhu, X., Huang, C., Tong, Y., Xu, B., Bai, J., Tong, J., et al. Spectral temporal graph neural network for multivariate time-series forecasting. *arXiv preprint arXiv:2103.07719*, 2021.

Chen, J., Xu, X., Wu, Y., and Zheng, H. Gc-lstm: Graph convolution embedded lstm for dynamic link prediction. *arXiv preprint arXiv:1812.04206*, 2018.

Chimmula, V. K. R. and Zhang, L. Time series forecasting of covid-19 transmission in canada using lstm networks. *Chaos, Solitons & Fractals*, 135:109864, 2020. ISSN 0960-0779.

Defferrard, M., Bresson, X., and Vandergheynst, P. Convolutional neural networks on graphs with fast localized spectral filtering. *arXiv preprint arXiv:1606.09375*, 2016.

Dong, E., Du, H., and Gardner, L. An interactive web-based dashboard to track covid-19 in real time. *The Lancet infectious diseases*, 20(5):533–534, 2020.

Dye, C., Cheng, R. C., Dagpunar, J. S., and Williams, B. G. The scale and dynamics of covid-19 epidemics across europe. *Royal Society open science*, 7(11):201726, 2020.

Facebook. Social Connectedness Index, 2021. URL https://dataforgood.fb.com/tools/social-connectedness-index/. Accessed: 8 May 2021.

Fauci, A. S., Lane, H. C., and Redfield, R. R. Covid-19 — navigating the uncharted. *New England Journal of Medicine*, 382(13):1268–1269, 2020.

Fritz, C., Dorigatti, E., and Rügamer, D. Combining graph neural networks and spatio-temporal disease models to predict covid-19 cases in germany. *arXiv preprint arXiv:2101.00661*, 2021.

Hamilton, W. L., Ying, R., and Leskovec, J. Inductive representation learning on large graphs. *arXiv preprint arXiv:1706.02216*, 2017.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Hochreiter, S. and Schmidhuber, J. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

La Gatta, V., Moscato, V., Postiglione, M., and Sperlí, G. An epidemiological neural network exploiting dynamic graph structured data applied to the covid-19 outbreak. *IEEE Transactions on Big Data*, 7(1):45–55, 2021.

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K. S., Lau, E. H., Wong, J. Y., Xing, X., Xiang, N., Wu, Y., Li, C., Chen, Q., Li, D., Liu, T., Zhao, J., Liu, M., Tu, W., Chen, C., Jin, L., Yang, R., Wang, Q., Zhou, S., Wang, R., Liu, H., Luo, Y., Liu, Y., Shao, G., Li, H., Tao, Z., Yang, Y., Deng, Z., Liu, B., Ma, Z., Zhang, Y., Shi, G., Lam, T. T., Wu, J. T., Gao, G. F., Cowling, B. J., Yang, B., Leung, G. M., and Feng, Z. Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia. *New England Journal of Medicine*, 382(13): 1199–1207, 2020.

Li, Y., Yu, R., Shahabi, C., and Liu, Y. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv preprint arXiv:1707.01926*, 2017.

Saglietto, A., D'Ascenzo, F., Zoccai, G. B., and De Ferrari, G. M. Covid-19 in europe: the italian lesson. *Lancet*, 395 (10230):1110–1111, 2020.

Santosh, K. Covid-19 prediction models and unexploited data. *Journal of medical systems*, 44(9):1–4, 2020.

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., and Monfardini, G. The graph neural network model. *IEEE transactions on neural networks*, 20(1):61–80, 2008.

Seo, Y., Defferrard, M., Vandergheynst, P., and Bresson, X. Structured sequence modeling with graph convolutional recurrent networks. In *International Conference on Neural Information Processing*, pp. 362–373. Springer, 2018.

Shahid, F., Zameer, A., and Muneeb, M. Predictions for covid-19 with deep learning models of lstm, gru and bi-lstm. *Chaos, Solitons & Fractals*, 140:110212, 2020. ISSN 0960-0779.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

United Nations. COVID-19 to slash global economic output by $8.5 trillion over next two years, 2020. URL https://www.un.org/development/desa/en/news/policy/wesp-mid-2020-report.html. Accessed: 8 May 2021.

Velavan, T. P. and Meyer, C. G. The covid-19 epidemic. *Tropical medicine & international health*, 25(3):278, 2020.

Wang, S.-H., Govindaraj, V. V., Górriz, J. M., Zhang, X., and Zhang, Y.-D. Covid-19 classification by fgcnet with deep feature fusion from graph convolutional network and convolutional neural network. *Information Fusion*, 67:208–229, 2021. ISSN 1566-2535.

World Health Organization. WHO Coronavirus (COVID-19) Dashboard, 2021. URL https://covid19.who.int/. Accessed: 8 May 2021.