How Bad Will Tomorrow Be? Covid Infection Rate Forecast using GRU

Samuel Londner
Tel-Aviv University
samuell@mail.tau.ac.il

Inbar Hasidim
Tel-Aviv University
inbarhasidim@mail.tau.ac.il

ABSTRACT

We leverage a small Gated Recurrent Unit (GRU) trained on publicly available data to perform multivariate regression and forecast the short-time spread of the disease in different towns in Israel. The originality of our research lays in extracting short-term features using a small GRU. The model attempts to infers relative intergeographic correlations. We find that a GRU has lower prediction error than current state-of-the-art models. Its performance is improved even more when incorporating a simpler model's prediction into the preprocessing stage. However, we find that the deep learning model gives comparable results to state-of-the-art method, and suggest that its potential may be better leveraged by fusing additional sources of data, such as encodings of the social distancing and lockdown regulations.

How Bad Will Tomorrow Be? Covid Infection Rate Forecast using GRU Availability Tag:

The source code of this research paper has been made publicly available at https://github.com/anutkk/covid19-israel/blob/main/Short_Term_Forecast_Using_GRU.ipynb.

1 INTRODUCTION

On 11 March 2020, the World Health Organization (WHO) declared the 2019 novel Coronavirus as a global pandemic. SARS-CoV-2, the etiological responsible for the disease COVID-19, originated in Wuhan, Hubei in China around December 2019. Since then, the virus has spread out all over the world, risking every living person and changing societies by causing public health issues, governmental restrictions, and economic instability. The first case in Israel was confirmed on 21 February 2020, when a female citizen tested positive for COVID-19 at the Sheba Medical Center after her return from quarantine on the Diamond Princess ship in Japan. Since then, the pandemic has largely spread in Israel, and currently (October 2020), Israel has one of the highest infection rates in the world.

To control the spread of the disease, Israel, like other countries in the world, imposed restrictions on its citizens to maintain social distancing and slow down the rate of infection. A model that would be able to predict the number of new cases without intervention would be helpful in preventing the spread of the disease by deciding how strict the limitations will need to be according to different parameters. Moreover, predictions per town would allow locally fine-tuning the limitations (this method having been colloquially nicknamed "traffic light program"). Moreover, such a model would be able to give a prediction of worst-case hospital load and other public health care services, so the public health system would be able to prepare accordingly. In our work we design a multivariate autoregressive model composed of a small Gated Recurrent Unit (GRU) which was trained on data made publicly available by the Israel Ministry of Health [8] to forecast the spread of the disease in different towns in Israel. Our network considers the impact of different towns on each other, and attempts to infer statistical geographic correlations.

There are multiple reasons for using GRU to forecast COVID-19 infection rate. First, COVID-19 cases are a time series data set and recurrent neural networks excel at extracting patterns from the data and performing nonparametric regression. Furthermore, many other prediction methods require various assumptions regarding the spread of the disease. For example, a statistical model can assume that the pandemic spread have a exponential or logistic shape in time and the method tries to find the parameters that best explains the data, although the data does not necessarily fit this gaussian model. One of the advantages of the GRU is that it finds a model that fit the data without any assumptions. It is highly relevant for the current pandemic, since COVID-19 is new and its epidemiological characteristics are not well-understood.

Our model is entirely autoregressive, i.e. based solely on past estimates of confirmed cases per inhabitants. The implicit assumption of such a model is a lack of change in epidemiological conditions, such as social distancing limitations or lockdown. Therefore its prediction is relevant in

order to simulate no intervention at all, and serve as a decision support tool for governmental intervention. Conversely, the model is appropriate for short-term predictions.

We find that in fact the GRU does *not* perform significantly better than state-of-the-art autoregressive models. This suggests that intergeographic influences are negligible in the case of Israel. The GRU potential may be better leveraged by using additional sources of data, like encodings of the social distancing and lockdown regulations or the number of deceased.

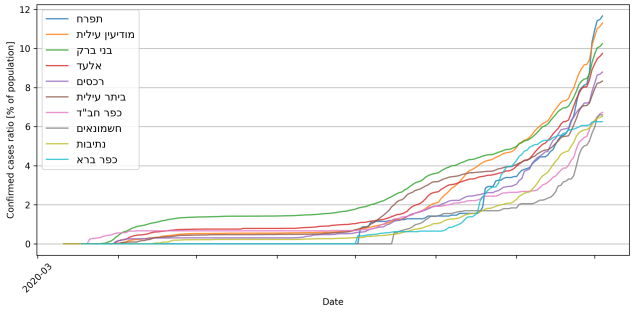


Figure 1: The confirmed case ratio for the ten towns with the highest case ratio

2 RELATED WORK

The use of deep learning for epidemiological prediction from a time-series perspective is relatively new. The first study on the topic was published in a seminal paper in 2018 [14]. The authors used an architecture combining RNN and Convolutional Neural Networks (CNN) to forecast weekly influenza-like-illness (ILI) activity levels. The purpose of the RNN is to capture the long-term correlation in the data and the CNNs to fuse information of data from different sources.

Further studies about more complex deep learning methods (like LSTM, GRU, and attention) try to take into account multiple factors like temperature, air pressure, medicine sales or even Google Trends (GT) data [3] [6]. Other studies attempt to design models that best forecast the spread of the disease in different countries. This kind of studies were done in Iran [6] and Canada [13]. In [5] several deep learning models are compared in order to fuse several features. Most of the studies compared their model with autoregressive (AR) models. In most researches the deep learning model provided consistently better results than these baseline methods.

Another notable example is a a 2019 research about forecasting Influenza-like Illness (ILI) using a small GRU [10]. The authors compare the performance of a GRU to other baseline statistic methods and machine learning methods, of real-time ILI estimations on the state and city level in the US. They found that the GRU is superior to baseline methods when there is a large reporting delay. Adding secondary data sources, like GT, did not improve meaningfully the model's performance. Finally, they conducted an in-depth analysis of feature importance for each model they build with the help of saliency maps.

It is of interest to note that research focusing on ILI generally uses several years of data and attempt to extract, at least in part, cyclic features of the series. This is obviously not possible with COVID-19, which has not yet exhibited such pronounced cyclic characteristics as ILI. As such, our model implicitly infers short-term epidemiologic parameters of the current spreading rate of the disease (such as Rt).

2.1 Our Contribution

We use a small GRU in order to extract short-term features of the spread of COVID-19, thus reducing greatly the complexity of the model compared to techniques adding CNN. We use a first-order estimate of the maximal value in order to normalize the data. We analyze the influence, among others, of the horizon and look-back window length hyperparameters. Furthermore, in order to simulate dynamic training of the model and real-time short-term prediction, the train and validation sets overlap by exactly the length of the look-back windows.

To the best of our knowledge, no research has attempted to extract autoregressive short-term characteristics using RNN.

3 DATA

We used the "COVID-19 By Area" dataset, published by the Israel Ministry of Health [8], to train and test the model. This dataset consists of the daily accumulated number of COVID-19 cases by "code town" (which may be smaller than a town) and by date. We aggregated the data and normalized to town population in order to get a multivariate time-series, where every feature is the confirmed cases ratio by resident in each town Figure 1. The dataset covers 262 towns between dates 11 March 2020 and 4 October 2020.

We also tried to apply two different scaling on the data. Applying scaling on the data can cause the gradient descent to converge much faster. Another reason for scaling is that if one of the features has a broad range of values, the loss will be governed by this particular feature. Therefore, the range of all features should be normalized so that each feature contributes approximately proportionately to the final distance.

One of our main contributions in this research is trying to find the right scaling that manages to achieve the best results. We applied two different scaling methods:

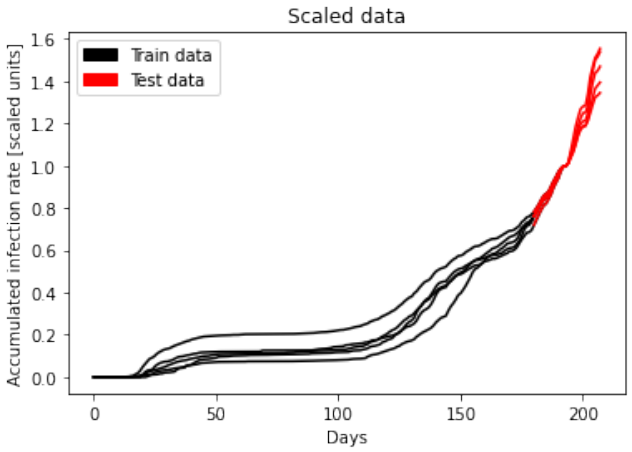


Figure 2: An example of 5 cities case rate after applying scaling. This graph shows a zero-one scaling of according to the train data.

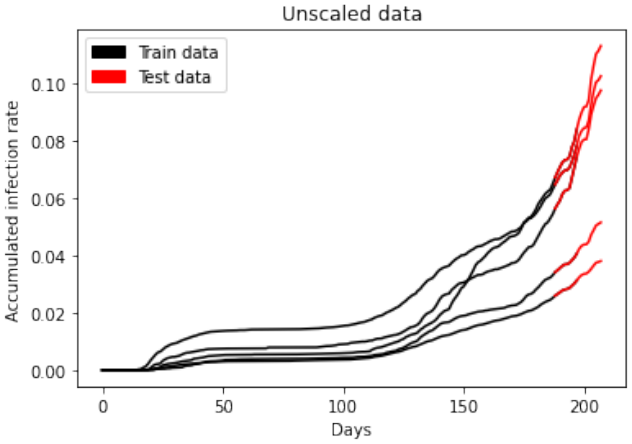


Figure 3: An example of 5 cities case rate before applying scaling.

- Normalization of the training data to be in the range of zero to one.
- First find a first-order estimator of the maximal value in the predicted values using the linear AR model. We then normalize all the data (training and validation sets) to be in the (estimated) range from zero to one.

We imported the data using the Ministry of Health API so the model can be easily updated. We also used a dataset of general information about Israeli towns, including population, from the CBS website [1]. This dataset contains the population of each town and we used it to normalize the number of cases by the number of inhabitants, resulting in

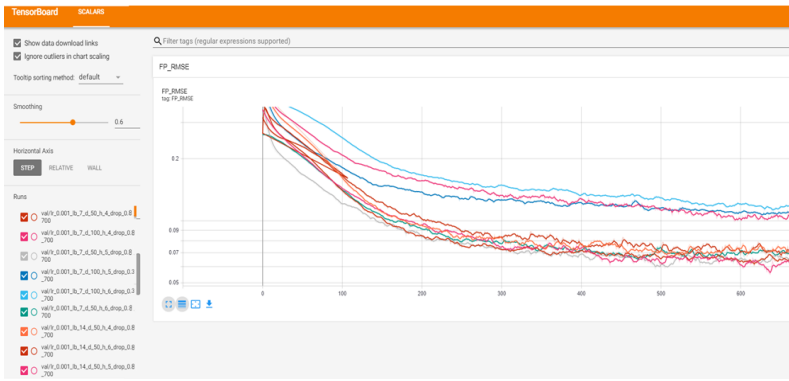


Figure 4: The error of the GRU model for different hyper-parameters.

infection rates for each town. The final results are displayed as absolute numbers of predicted confirmed cases.

4 METHODS

The classical forecasting models mainly use linear regression [15] [7] or, less often, random forests or support vector machines [11]. However, the complexity of epidemiological models calls for more complex nonparametric models. Here we consider the usefulness of RNN, which are well-suited to perform nonparametric regression on time series data [9] [4].

Following [10], we implement a small Gated Recurrent Unit Neural Network (GRU) with a single hidden layer. The GRU performance is generally on par with its cousin alternative LSTM, and is computationally more efficient [2]. Since we aim to make short-term predictions (up to 14 days), we built an entirely autoregressive model. We note that the model may also give reasonable long-term prediction, assuming there is no change in epidemiological conditions.

4.1 GRU

Gated recurrent units (GRUs) are a gating mechanism in recurrent neural networks, introduced in 2014 by Kyunghyun Cho et al [12] The GRU is similar to a long short-term memory (LSTM) with a forget gate, but has fewer parameters than LSTM, as it lacks an output gate.The model can be described mathematically in the following way:

$$r_t = \sigma\left(W_{ir}x_t + b_{ir} + W_{hr}h_{t-1} + b_{hr}\right)$$
$$z_t = \sigma\left(W_{iz}x_t + b_{iz} + W_{hz}h_{t-1} + b_{hz}\right)$$
$$n_t = tanh\left(W_{in}x_t + b_{in} + r_t * \left(W_{hn}h_{t-1} + b_{hn}\right)\right)$$
$$h_t = (1 - z_t) * n_t + z_y * h_{t-1}$$

where h_t is the hidden state at time t, x_t is the input at time t, h_{t-1} is the hidden state of the layer at time $t-1$ or the initial hidden state at time 0, and r_t, z_t, n_t are the reset, update and

new gates respectively. σ is the sigmoid function and * is the Hadamard product.

The GRU accepts as input 7 autoregressive terms from all locations in the data set and predicts incidence at the given time horizon for all locations simultaneously. The GRU is trained on a mean-squared error objective with a dropout rate of 0.7 after the hidden layer to reduce overfitting. We use a learning rate of 0.001 for stochastic gradient descent and train the model for 1000 epochs.

4.2 Baseline Models

To examine the results of the GRU model, we compared its performance to the results of three baseline models.

4.2.1 Persistence model. :

The persistence (**P**) model is the standard naïve baseline for time-series prediction, in which the most recently observed incidence is propagated h weeks forward.

4.2.2 Linear Autoregression. :

The linear autoregression (**AR**) uses a linear combination of N autoregressive observations of COVID cases in a given town to predict incidence at time horizon h in that location. The coefficients are obtained using ordinary least squares (OLS) optimization. This means that the model has $c \cdot l$, where l is the size of the window length and c is the number of towns.

4.2.3 Linear Network Autoregression. The linear network autoregression (**LR**) captures spatial spread of disease, taking as features a linear combination of N autoregressive terms from a set of regions R available in the data set.

4.3 Experiments

All models are evaluated on the test set based on the root mean squared error (RMSE) across all locations and dates.

Table 1: The RMSE of the models

Model	RMSE
Persistence model	4.14
Auto Regression model	0.87
The linear network autoregression	2.12
GRU	2.03

4.4 Baseline Models

As we can see from table **??** in the appendix, the persistent model has the highest error. That make sense since this model is too simplistic and it clearly not sophisticated enough to describe the evolving data.

One can also assume that maybe the linear network auto regression is a too complicated model. This model assumes that the spread of the pandemic in different towns affect each other. The model tries to find the correlation between the cities and how they affect one another. This correlation doesn't necessarily exist and maybe forcing the model to find it will cause to overfitting of the data. We will come back to this point when we discuss the hyper-parameters of the GRU model.

As we assumed, the models easily fits the true data when we try to predict the next three or four days. The mission of prediction becomes harder when the models tries to predict the rate of cases in the far future. The long-term features may be dependant on additional data and autoregressive models may not be powerful enough to extract them . This will be discussed in the concluding section.

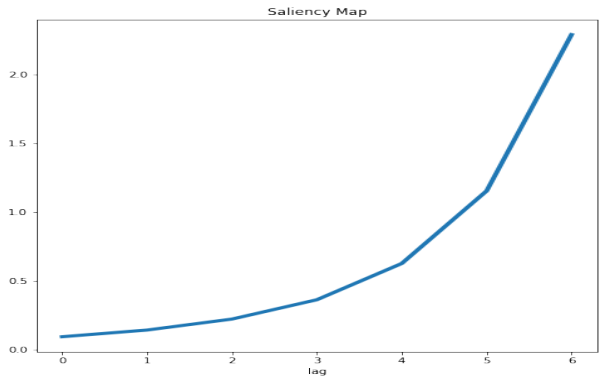


Figure 5: Saliency map for Jerusalem

4.5 GRU

In order to simulate real-life conditions and dynamic training, the models were trained on all the series, excluding the last h samples, h being the horizon of the model. The models are evaluated on the last $h + w$ samples, where w is the size of the lookback window.

The sample saliency map in Figure 5 how each day in the 7 last days influences the prediction. As expected, the last days affects the prediction the most. This is expected for short-terms predictions.

The other hyper-parameters we considered were: number of epochs, look-back window size, dropout rate, size of hidden layers, number of towns and the normalization.

We considered how many towns from the data set to give to the model. We find that giving too many towns actually decreases the performance. This seems to be because the model tries to find correlations between towns wheres they are in fact weakly correlated.

From Figure 4 we can see that the dropout should be 0.7 (all the experiments with the lower dropout achieved higher

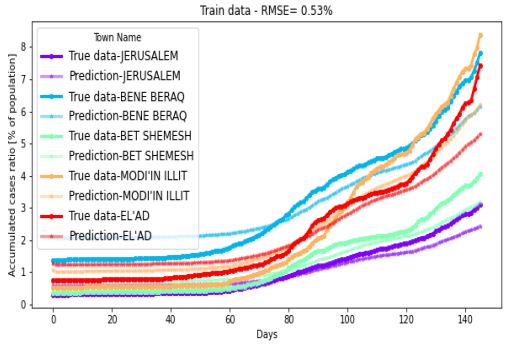


Figure 6: Train set - GT and predictions

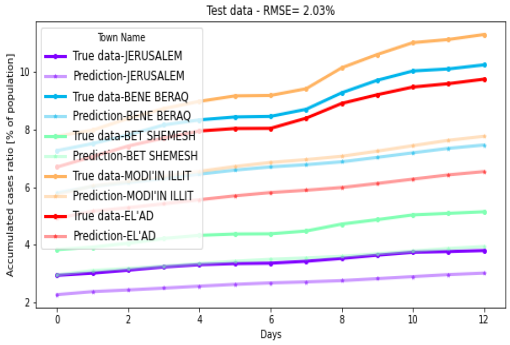


Figure 7: Validation set - GT and predictions

validation error). We can also see that the model starts to overfit the data around epoch 1000. Another thing that we can see from this graph is that the performance is not influenced if the look-back window length is 7 or 14. This suggests the parameters of the pandemic spread changes slowly, even when restrictions change. Therefore, the parameters extracted from 7 days are similar to the parameters extracted from 14 days of data.

We examined different learning rates and find out that the value 0.001 achieves good results in reasonable time.

From these graphs and from other experiment that we did we choose the parameters which yield the best validation error. The parameters that where chosen are:

- Number of epochs: 1000
- Look-back window length: 7
- Dropout rate: 0.7
- Size of hidden layer: 4
- Number of towns: 5

Figure 6 and Figure 7 show the performance of the model on the train and validation sets. As expected, the error on the training set is a slightly lower than the error on the validation set. The model succeeds in approximately predicting the number of cases in the next few days and it is doing it much better than the simple linear NN model and the persistent model.

The performance of the GRU is comparable to the linear AR model. Although the GRU has more parameters, this seems to be because the GRU attempts to find correlations between towns. In practice this assumption does not seem to be accurate, which puts the AR model at an advantage. When trained over every city separately, the GRU attains similar RMSE (Figure 8)

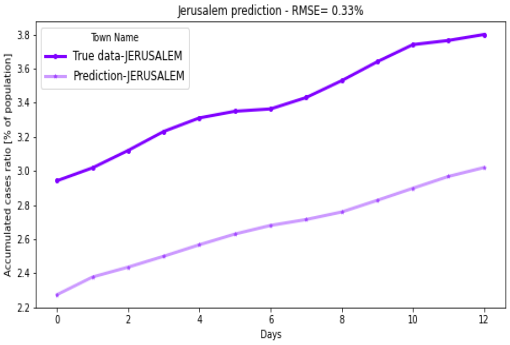


Figure 8: Sample performance of a GRU trained on a single town.

5 CONCLUSION

In this research we examine the performance of the GRU model with comparison to classic models, in forecasting the number of COVID-19 cases at the town level in Israel. We find that although the GRU gives slightly better results, its performance is not significantly better. We suggest this happens because RNN were designed to learn long-term features, and do not bring a significant advantage in short-term prediction. Therefore, our model gives relatively accurate short-term predictions, but not necessarily better than state-of-the-art models. On the other, The size and lack of stability of the data (due to the changing number of daily COVID-19 tests) may also have undermined the GRU's advantage, which on the long term diverges from the true numbers.

We also find that the model's performance is insensitive to the size of the look-back window, which leads to the conclusion that the underlying epidemiological model evolves slowly, and a few datapoints are enough to extract meaningful features for short-term prediction.

In order to give better long-term predictions and leverage the power of the GRU, we suggest to add as features the number of deceased from COVID-19 (which has also been made available by the Ministry of Health), as well as a one-hot encoding of the social distancing and lockdown regulations in

place along time. To the best of our knowledge, this dataset does not exist yet and would have to be built, mainly from online sources.

ACKNOWLEDGMENTS

This work was made as a requirement for the course "Deep Learning", by Dr. Raja Giryes, and was supported by Google Cloud Platform credits provided by the course's staff.

REFERENCES

[1] Central Bureau of Statistics [n.d.]. *General information about Israeli towns.* Central Bureau of Statistics. Retrieved 18 May, 2020 from https://www.cbs.gov.il/he/publications/doclib/2019/ishuvim/bycode2019.xlsx

[2] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. 2014. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. (2014). arXiv:1412.3555 [cs.NE]

[3] Bofeng F. Yaodong Y. Jianye H. 2019. Attention-based recurrent neural network for influenza epidemic prediction. (Nov. 2019), ,. https://doi.org/,

[4] S. L. Ho, M. Xie, and T. N. Goh. 2002. A Comparative Study of Neural Network and Box-Jenkins ARIMA Modeling in Time Series Prediction. *Comput. Ind. Eng.* 42, 2–4 (June 2002), 371–375. https://doi.org/10.1016/S0360-8352(02)00036-0

[5] Chiou-Jye Huang, Yung-Hsiang Chen, Yuxuan Ma, and Ping-Huan Kuo. 2020. Multiple-Input Deep Convolutional Neural Network Model for COVID-19 Forecasting in China. *medRxiv* (2020). https://doi.org/10.1101/2020.03.23.20041608

[6] Seyed M. Ayyoubzadeh S. Mehdi A. Hoda Z. Mahnaz A. Sharareh R. Niakan K. 2020. Predicting COVID-19 Incidence Using Google Trend and Data Mining Techniques: A case study of Iran. (March 2020), ,. https://doi.org/10.2196/preprints.18828

[7] Fred Lu, Mohammad Hattab, Cesar Clemente, Matthew Biggerstaff, and Mauricio Santillana. 2019. Improved state-level influenza nowcasting in the United States leveraging Internet-based data and network approaches. *Nature Communications* 10 (12 2019). https://doi.org/10.1038/s41467-018-08082-0

[8] Israel Ministry of Health. [n.d.]. *covid-19 by area.* Israel Ministry of Health. Retrieved May 27, 2019 from https://data.gov.il/dataset/covid-19/resource/d07c0771-01a8-43b2-96cc-c6154e7fa9bd

[9] Yao Qin, Dongjin Song, Haifeng Chen, Wei Cheng, Guofei Jiang, and Garrison Cottrell. 2017. A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction. (2017). arXiv:1704.02971 [cs.LG]

[10] Emily L. Aiken Andre T. Nguyen Mauricio Santillana. 2019. Towards the Use of Neural Networks for Influenza Prediction at Multiple Spatial Resolutions. (Nov. 2019), 36–44. https://doi.org/1911.02673

[11] Mauricio Santillana, Andre Nguyen, Mark Dredze, Michael Paul, and John Brownstein. 2015. Combining Search, Social Media, and Traditional Data Sources to Improve Influenza Surveillance. *PLoS computational biology* 11 (08 2015). https://doi.org/10.1371/journal.pcbi.1004513

[12] Kyunghyun C. Bart van M. Caglar G. Dzmitry B. Fethi B. Holger S. Yoshua B. 2014. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. (June 2014), ,. https://doi.org/1406.1078

[13] Lei Zhang Vinay Kumar Reddy Chimmula. 2020. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. (May 2020), ,. https://doi.org/10.1016/j.chaos.2020.109864

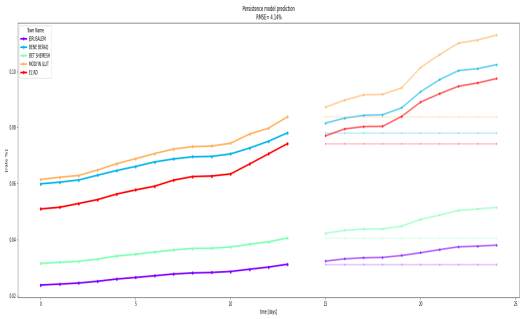
[14] Yuexin Wu, Yiming Yang, Hiroshi Nishiura, and Masaya Saitoh. 2018. Deep Learning for Epidemiological Predictions. 1085–1088. https://doi.org/10.1145/3209978.3210077

[15] Shihao Yang, Mauricio Santillana, and S. C. Kou. 2015. Accurate estimation of influenza epidemics using Google search data via ARGO. *Proceedings of the National Academy of Sciences* 112, 47 (2015), 14473–14478. https://doi.org/10.1073/pnas.1515373112 arXiv:https://www.pnas.org/content/112/47/14473.full.pdf

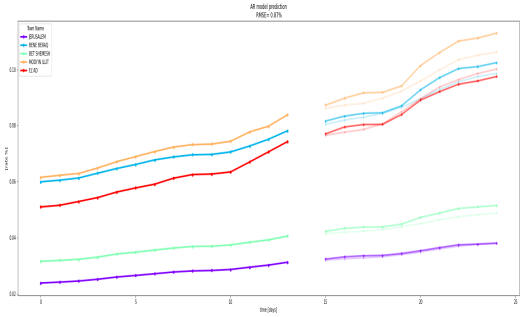
Appendices

A THE PREDICTIONS OF THE NAIVE MODELS

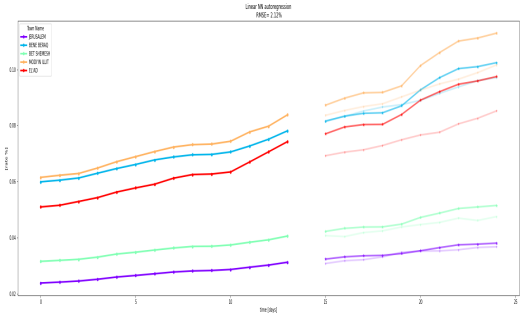
Model	Graph



Persistence model



AR model



Linear NN

Table 2: The matte lines are the true data that was taken from the ministry of health database. The bright plots are the predictions of the different naive models.