PREPARING FOR DISEASE X: PREDICTING ICU ADMISSIONS USING TIME SERIES FORECASTING WITH DECODER-ONLY TRANSFORMER NEURAL NETWORKS

NEJC ČELIK, ANDREJ ŠKRABA
University of Maribor, Faculty of Organizational Sciences, Cybernetics & Decision Support Systems Laboratory, Kranj, Slovenia
nejc.celik1@um.si, andrej.skraba@um.si

The COVID-19 pandemic has underscored the critical importance of predictive modelling in managing healthcare resources and shaping public health policies. This paper explores the application of advanced Artificial Intelligence (AI) techniques, specifically decoder-only transformer neural networks (DOTNN), in forecasting weekly Intensive Care Unit (ICU) admissions. Our research is driven by the necessity to enhance preparedness for potential future pandemics, referred to as "Disease X", by leveraging large datasets of publicly available information. A prediction model has been developed that incorporates several key indicators, such as new cases, ICU admissions, and testing rates. Our DOTNN architecture, inspired by the Generative Pre-trained Transformer (GPT), focuses on time series forecasting without the necessity for encoder components, thereby streamlining the prediction process. Despite limited data availability, the proposed method can achieve notable accuracy, with Mean Absolute Percentage Error (MAPE) values below 15% for a significant number of predictions. This performance highlights the potential of DOTNNs in forecasting ICU admissions, which is crucial for healthcare planning and resource allocation during pandemics.

Keywords: ICU admission predictions, decoder-only transformer neural networks, disease X, pandemic preparedness, COVID-19 forecasting

University of Maribor Press
DOI https://doi.org/10.18690/um.fov.4.2024.43
ISBN 978-961-286-871-0

1 Introduction

The COVID-19 pandemic spurred a wide range of modelling efforts to predict the spread of the virus, inform public health policies, and manage healthcare resources. While numerous methodologies were employed, many of these efforts were indeed grounded in classical epidemiological models like the SIR (Susceptible-Infected-Removed), SEIR (Susceptible-Exposed-Infected-Removed), SEIUR (Susceptible-Exposed-Infected-Uncertain-Removed), and agent-based models. These models have a long history in epidemiology for their utility in understanding the dynamics of infectious diseases. In contrast, advanced Artificial Intelligence (AI) methods constituted a smaller proportion of the research landscape during the initial stages of the pandemic (Adam, 2020; Kucharski et al., 2020).

In the aftermath of the COVID-19 era, one might question the need to continue refining pandemic prediction methods. However, it's essential to consider the concept of "Disease X"—a potential unknown threat that underscores the critical importance of preparing for yet-undiscovered pathogens capable of sparking future pandemics. Proactive stance in research of prediction methods for pandemics can markedly improve prepares for future pandemics. Such forward-thinking strategies are vital for saving lives, preventing the recurrence of past errors, and ensuring an effective response to novel infectious threats. (Banerjee et al., 2023; *Prioritizing Diseases for Research and Development in Emergency Contexts*, n.d.; *What Is Disease X | Johns Hopkins | Bloomberg School of Public Health*, n.d.)

Our study aims to apply cutting-edge AI methodologies to a vast dataset for the purpose of forecasting ICU admissions. Rather than refining prediction algorithms, we seek to evaluate the efficacy of employing transformer neural networks for this task. Specifically, we are interested in predicting weekly ICU patient counts based on historical data, mirroring real-world scenarios where daily infection rates are recorded. This predictive capability holds promise for optimizing resource allocation and preparedness in managing COVID-19 patients. While existing research predominantly focuses on predicting ICU requirements for confirmed cases (Lorenzen et al., 2021) (Subudhi et al., 2021) (Chadaga et al., 2024) (Dipaola et al., 2023), our methodology adopts a novel approach, utilizing time series forecasting with readily accessible data and a customized transformer architecture. However, our analysis is constrained by the limited availability of comprehensive weekly ICU

admission data from fewer than 20 countries, highlighting the necessity for enhanced data acquisition protocols, particularly in anticipation of future outbreaks. Despite potential challenges arising from data scarcity, precise forecasting of ICU occupancy remains pivotal for ensuring sufficient capacity to accommodate all critically ill patients.

Development of Covid-19 models or "digital twins" should incorporate the real time sensing. In our early efforts in this regard, we have developed the hardware interface (Stojanovic et al., 2020) to monitor Covid-19 patients and provide the input to the predictive simulation models. In the next stage of our previous attempts to simulate the Covid-19 spread we have applied Bass diffusion model (Škraba et al., 2021). Agent-based approach has also been applied (Škraba & Vavtar, 2022) addressing the spatial distribution of the infected population. In order to improve the accuracy of the predictions the set of models was used: SI, Bass diffusion, SIR, SEIR and SEIUR (Stanovov et al., 2022) Parametrization was performed on parallel computer stack with the Differential Evolution (DE) methods. Combination of standard models with DE enabled us to confirm the hypothesis of latent spread mechanisms of Covid-19 which are important for the epidemics prediction. In present paper we strive to extend the predictability accuracy with the transformer neural network model.

2 Methodology

2.1 Data 2.1

To develop our prediction model, we used publicly available dataset on COVID-19 infections (*Data on COVID-19 (Coronavirus) by Our World in Data*, n.d.) which includes data of number of confirmed cases, tests, intensive care units (ICU) admissions, number of hospitalizations and more per country. Data for number of weekly hospitalizations and intensive care patients for at least 102 weeks is only available for 17 countries. For input data of our model, we used some of the data from dataset (new_cases_smoothed_per_million, weekly_icu_admissions_per_million, date (as number YYMMDD), aged_65_older (fixed per country), aged_70_older (fixed per country), gdp_per_capita (fixed per country), cardiovasc_death_rate (fixed per country), diabetes_prevalence (fixed per country), female_smokers (fixed per country), male_smokers (fixed per country), population (in millions) (fixed per

country), weekly_hosp_admissions_per_million (updated weekly), population_density (fixed per country), median_age (fixed per country), life_expectancy (fixed per country), human_development_index (fixed per country), new_deaths_smoothed_per_million(updated weekly), new_tests_per_thousand (updated weekly)).

Figure 1 shows five time series per country that were used in the training phase. Here the timeseries for Chile (CHL) are shown as an example in order to illustrate the complexity of addressed task. The graph includes the following time series data:

- new cases, smoothed per million people, which exhibits several peaks that correspond to waves of infections over time.
- weekly intensive care unit (ICU) admissions per million people, which shows notable spikes that are usually correlated with the waves of new cases.
- new deaths, smoothed per million people, again showing peaks which typically follow the trend in new cases with a certain lag.
- weekly hospital admissions per million people, closely following the trends of the ICU admissions, but typically at higher levels, suggesting that not all hospital admissions require ICU care.
- new tests conducted per thousand people which has visible drops on weekends, highlighting the consistency of testing throughout the pandemic except for lower testing rates on weekends.

To represent the dynamic character of the input data, the timeseries were normalized on the interval [0, 1] (y-axis), which makes it easier to compare the different scales of data. The normalization process transformed the raw data so that each metric fits within the same range for comparative purposes. The "waves" of the pandemic are evident from the recurring peaks in new cases, hospitalizations, ICU admissions, and deaths, with testing showing periodicity Figure 1 shows the time series for only one of the seventeen countries which were used to train the neural network models.

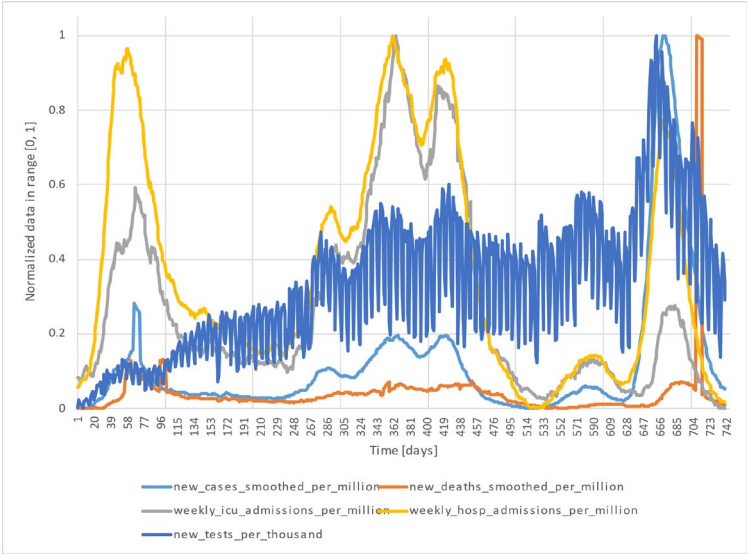

Figure 1: Normalized time series that were used as the input for Chile (CHL)
Source: Own

2.1.1 Model 2.1.1

Our architecture leverages a decoder-only transformer neural network (DOTNN) model inspired by GPT (Generative Pre-trained Transformer) as detailed by Radford et al. (2018). This design choice is a deviation from the original transformer architecture proposed by Vaswani et al. (2017), focusing on sequence generation task without the need for an encoder. This decision stems from the nature of our forecasting task, which primarily involves generating future values from past data rather than transforming one sequence into another.

Following the methodology outlined by Wen et al. (2022) and Zhou et al. (2020), we implement zero-mean normalization on both model input and target data... During inference, we apply the inverse process (de-normalization) using the training data's mean and variance to transform model outputs back to their original scale, ensuring that our predictions are interpretable and comparable to real-world values. In our

model we employ learned positional embeddings (Wen et al., 2022; Zhou et al., 2020).

In our model before decoder blocks, we expand inputs dimensionality to 512 and combine it with positional embeddings, each decoder block is equipped with two masked multi-head attention layers featuring four attention heads. The attention mechanism is complemented by a feed-forward network with 512 neurons. As we started predictions after 20 weeks context length of 20 weeks was chosen and was then kept through all subsequent weeks. We have also developed a smaller model with only one decoder block and only one attention head to expedite the experimental phase since model training takes considerable computer time.

For training, we utilized a cosine learning rate decay strategy (Loshchilov & Hutter, 2016), starting with an initial rate of 3e-4 and gradually reducing it to 1e-4 across the epochs. Coupled with the AdamW optimizer, which introduces a decay component to the weights (Kingma & Ba, 2014) (Loshchilov & Hutter, 2017) All examples were feed into network in a single batch (1 step per epoch).

Figure 2 illustrates our network architecture, adapted from Vaswani et al. (2017), detailing the composition and workflow of our model.

Our experimental setup involves training the model on a dataset covering 20 weeks of data, aiming to predict ICU patient numbers in the subsequent week. This process is iteratively repeated, incorporating the data from the newly predicted week into the training set for subsequent predictions. This iterative retraining strategy simulates a real-world application where the model adapts to new information over time, aiming to provide accurate forecasts. During development we used up to 102 weeks of data.

To accommodate the growing dataset, we proportionally increase the number of training epochs based on the added examples, opting for a pragmatic approach over hyperparameter optimization for each iteration.

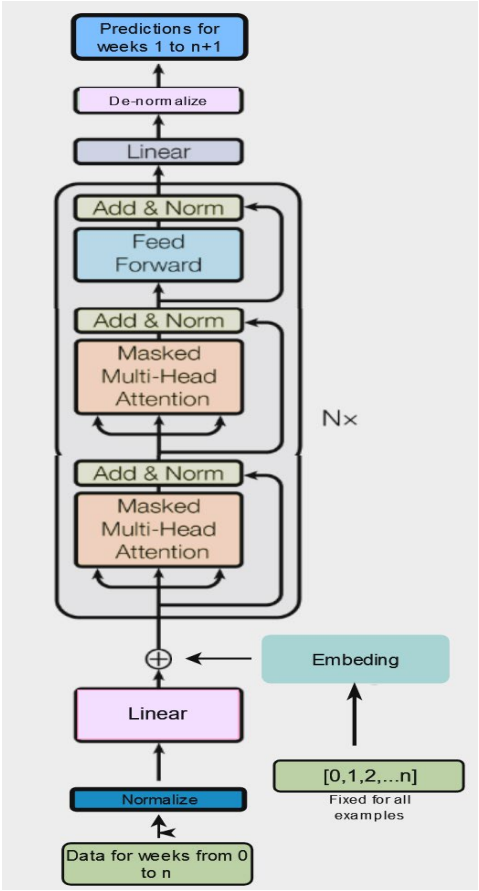

Figure 2: Network architecture
Source: adapted from (Vaswani et al., 2017)

As the measure to determine the accuracy of a model's predictions the Mean Absolute Percentage Error (MAPE) (Oliva & Oliva, 1995; Sterman, 2000) was used, defined as:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \times 100\%$$

where n is the number of observations, $\hat{y}_i$ is the predicted value and y_i is the actual value. While, due to the absolute value, the term $\hat{y}_i - y_i$ is sometimes swapped which might contribute to the confusion, the proper order as is written enables us to understand the meaning of MAPE. There is also plethora of other possibilities to determine the accuracy of the predictions (Batagelj & Bren, 1995) however, we have used MAPE due to its intuitive interpretability, as it directly expresses error as a percentage of the actual values.

3 Results

With proposed method of using DOTNN for predicting the number of ICU patients the MAPE values for 6 out of the 17 countries were below 15% which might be considered good in terms of ICU prediction. Observing the results we saw that in some countries weekly ICU admissions were logged daily for those countries our model made daily predictions (Table 1, marked with asterisk *) and performed better in general. We have tested our method with two different model sizes. While the larger model exhibited a superior average MAPE, the smaller model demonstrated better performance in countries where the MAPE was below 15%.

Bolded values in Table 1 indicates better, i.e. lower MAPE values indicating better model. The same goes for the MSE (mean squared error) while for the r2 (Wright, 1921) higher values are better. Asterisk * marks the countries where daily data was used for training.

In examining the accuracy of predictions, t-test was conducted to compare the mean accuracy of predictions with a MAPE below 25% (M = 8.70, SD = 3.09) against those with a MAPE above 25% (M = 87.81, SD = 58.45). Results indicated a statistically significant difference between the two groups, $t(11) = -4.68$, $p < .001$, two-tailed. The significant statistical difference in prediction accuracy, as evidenced by the t-test comparing MAPE values below 25% to those above this threshold, highlights a potential for enhancing the prediction method as well as good accuracy in the <25% group.

Table 1: Prediction scores

ISO CODE	MAPE (bigger model)	MSE	r2	MAPE (smaller model)	MSE	r2
CHL*	7.36	8.31	0.90	**4.69**	**5.56**	**0.93**
CYP	**64.78**	**31.69**	**0.60**	113.31	38.11	0.55
CZE*	**112.03**	0.36	0.84	152.98	**0.25**	**0.88**
EST	**52.51**	**45.32**	**0.70**	112.88	52.16	0.53
FRA*	**11.44**	**0.51**	0.99	13.00	0.52	0.99
DEU*	12.77	2.03	0.82	**5.50**	**0.32**	**0.98**
GRC	34.36	**16.23**	0.85	**30.28**	17.38	**0.86**
IRL	**137.52**	19.35	**0.35**	230.85	**16.72**	0.12
ISR*	244.66	1.21	0.84	**191.91**	**0.72**	**0.90**
ITA*	7.59	4.22	0.64	**4.97**	**2.10**	**0.86**
LVA	50.42	**380.02**	**0.73**	**49.15**	697.01	0.43
LUX	**75.45**	**46.59**	0.59	84.28	57.49	**0.60**
NLD*	8.86	2.40	0.80	**7.23**	**1.80**	**0.85**
NOR	**79.01**	**1.94**	**0.77**	246.66	2.81	0.72
SVK	**105.28**	**13.17**	0.81	163.89	14.27	**0.82**
SVN	**36.30**	**73.48**	**0.81**	53.57	93.41	0.76
ESP*	**4.16**	**1.21**	0.98	4.84	1.41	0.98
AVERAGE	**61.44**	**38.12**	**0.77**	86.47	58.94	0.75

Figure 3 represents the overall best forecasting accuracy for ICU admissions which is for Spain (ESP) over a span of 102 days, using a daily scale. The blue line signifies the actual reported ICU admissions per million individuals, while the red arrows illustrate our model's daily predictions. The accurate alignment between the blue line and the red arrows demonstrates the high precision of the predictions, with the MAPE being exceptionally low at 4.16%, indicating a close correspondence to the real data. The model's capacity to closely follow the trend lines throughout the entire

time highlights its robustness and the potential for reliable future forecasting in similar scenarios.

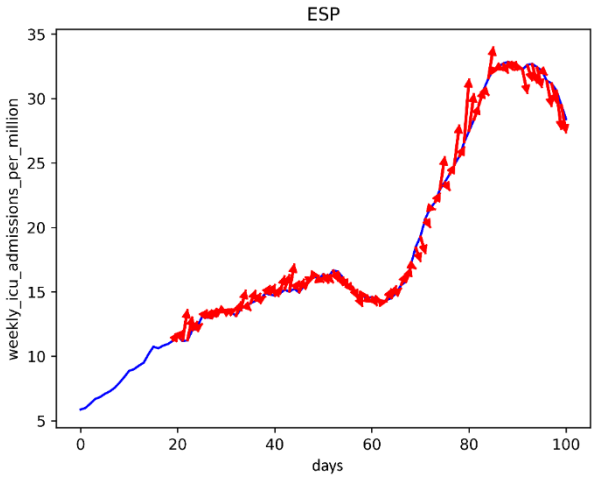

Figure 3: Best performing country (ESP bigger model)
Source: Own

Figure 4 presents a comparative analysis of ICU admission predictions for the three countries where both the smaller model (left column) and the bigger, more accurate model (right column) performed the least accurately. Each row represents one country, with Israel (ISR), Norway (NOR), and Ireland (IRL) from top to bottom. The time series data spans 102 days/weeks and shows weekly ICU admissions per million. The blue lines indicate the actual data for ICU admissions, while the red lines depict the predictions made by the respective model.

Figure 4 reveals the limitations and challenges faced by the models, particularly in the countries where predictions did not align as closely with the actual data. Some of the error can be explained by extreme errors when predictions are close to 0 which is known disadvantage of MAPE. We observed the smaller model does not adequately account for all the information contained in the data. As the model gets larger, the predictions in our case improved on average. Model might be also missing some key data for making best predictions as for example amount and strictness of COVID measures and restrictions and country's location and size.

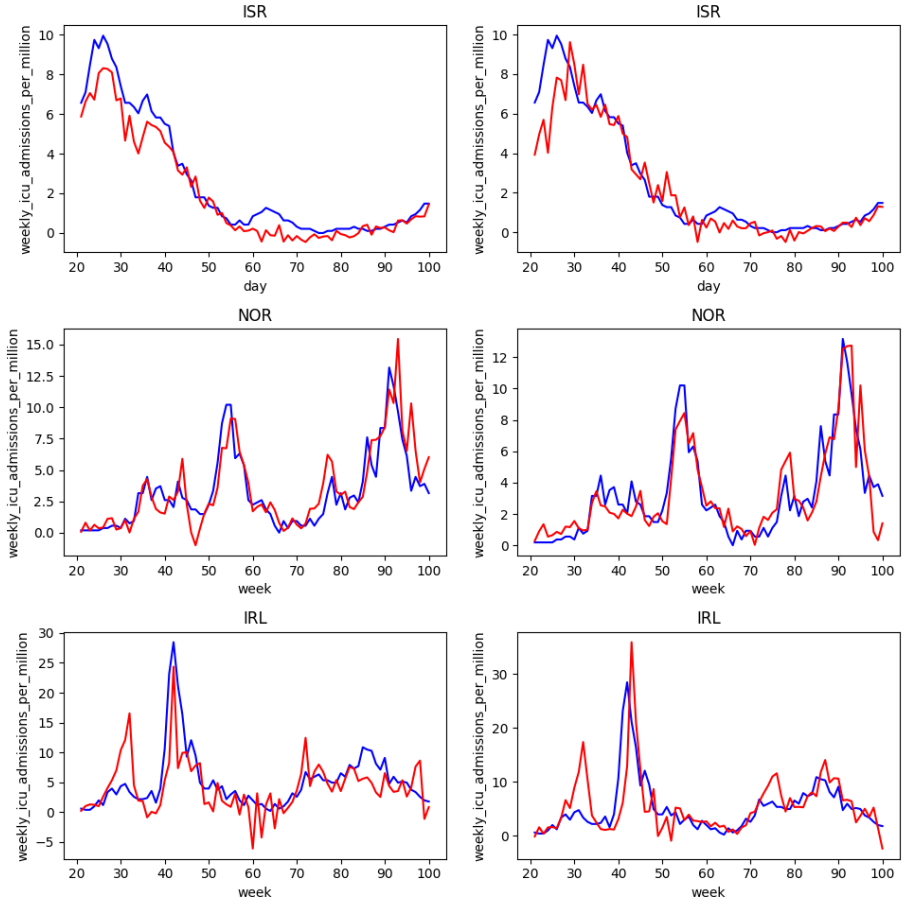

Figure 4: worst performing countries (left smaller model, right bigger model)
Source: Own

To explore relationship between amount of data and prediction accuracy we selected the smaller set of countries to compare the results between smaller and larger data set. In the left part of table 2 (smaller model separate data), we have only one country in the training set and not all of seventeen selected countries. The set of input data is correspondingly smaller in this part. The right side of the table (smaller model all data) shows the results with all countries in the training set.

Table 2: Results for smaller (left) vs bigger (right) training set

ISO CODE	MAPE	MSE	r2	MAPE	MSE	r2
GRC	48.54	24.04	0.80	**30.28**	**17.38**	**0.86**
NOR	**74.83**	5.02	0.40	246.66	**2.81**	**0.72**
SVK	**77.70**	16.90	0.77	163.89	**14.27**	**0.82**
SVN	327.03	**81.97**	**0.82**	**53.57**	93.41	0.76
ITA	**3.15**	**0.75**	**0.96**	4.97	2.10	0.86

In the case of smoother data input, where there was small volatility in data the model trained on less data turns out to be better. With harder to predict countries we observed better performance if we included all countries in the training dataset. Considering better MSE and r2 scores for Norway and Slovakia predictions one could hypothesise that better MAPE might be explained by lower number of extreme errors when predictions are near zero. Figure 5 ~ Left shows predictions for Greece when trained with only Greece's data and Figure 5 ~ Right shows predictions when trained on all data. Both cases used smaller model for prediction.

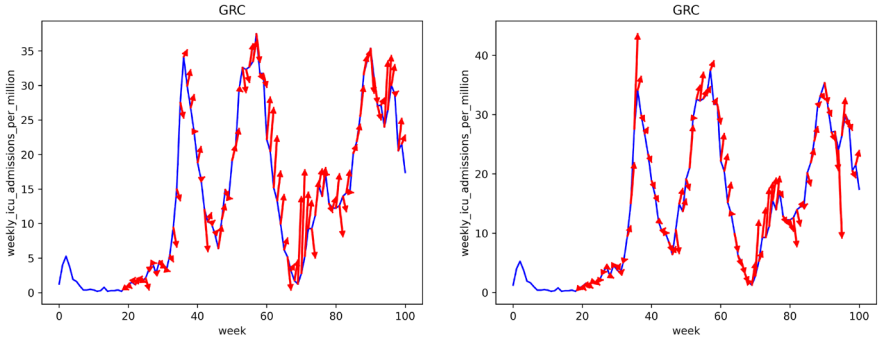

Figure 5 Left ~ GRC trained only on GRC data | Right ~ GRC trained on all data (smaller model)

Comparatively, the data in Figure 5 ~ Right may represent a more stable and possibly more accurate model for predicting ICU admissions, incorporating a diverse set of training data. However, Figure 5 ~ Left, with its higher resolution of Greece-specific fluctuations, might be more sensitive to local variations and potentially overfit to

Greece's patterns. With better performance observed with more data and a bigger model, gathering detailed data for regions or municipalities could potentially increase accuracy of predictions while adding additional value for planning. Pretraining the model on different diseases is also worth exploring.

Main possible advantage of proposed method of predicting number of ICU admissions is the ability to put any data as an input and then let the model to learn if the data is valuable to making accurate predictions or if it should be ignored. Additionally, adding additional inputs to the model does not increase the complexity of modelling. Previous methods such us Google Flu Trends (Dugas et al., 2013)and Skraba (Škraba & Vavtar, 2022) were based on rather different principles as one was monitoring user searches and behaviours in Google search tools and the other was predicting epidemic dynamics for a single wave of pandemic. Our methodology could also provide ability to easily include predictions of other methods as inputs with zero additional modelling complexity increase besides gathering the data.

4 Conclusion

We've shown that the DOTNN method could offer high accuracy in predicting ICU admissions numbers, though occasional higher prediction errors can occur. Future research should focus on data preparation and identifying factors influencing inaccurate predictions. Despite some discrepancies, many predictions were satisfactory, with MAPE below 15%. Our aim wasn't to dive into prediction accuracy but to explore a novel method, unused in the context of COVID-19. This approach prepares for potential future pandemics, like "Disease X," where AI methods could be pivotal. While our methodology shows promise, its usability remains unclear due to data limitations. We observed better accuracy predicting ICU admissions a day ahead compared to a week ahead which indicates that predicting for less than week in advance may be worth exploring. We believe that for future research data preparation is critical, with need for manual inspection and filtering due to potential errors. Adapting these methods for real-time processing is vital, particularly for future pandemics. Despite challenges in modelling pandemic waves, our method shows potential, although comparing it with others is difficult due to data complexity.

Acknowledgements

The authors acknowledge the financial support from the Slovenian Research and Innovation Agency (research core funding No. P5-0018) and the Ministry of Higher Education, Science, and Innovation of the Republic of Slovenia as part of the Next Generation EU National Recovery and Resilience Plan (Grants No. 3330-22-3515; NOO No: C3330-22-953012).

References

Adam, D. (2020). Special report: The simulations driving the world's response to COVID-19. Nature, 580(7803), 316–318. https://doi.org/10.1038/D41586-020-01003-6

Banerjee, S., Sultana, Q., Mukherjee, D., Agrawal, V., Jaiswal, V., & Paudel, K. (2023). Disease-X: Accounting for the unknown. Health Science Reports, 6(4). https://doi.org/10.1002/HSR2.1173

Batagelj, V., & Bren, M. (1995). Comparing resemblance measures. Journal of Classification, 12(1), 73–90. https://doi.org/10.1007/BF01202268/METRICS

Chadaga, K., Prabhu, S., Sampathila, N., Chadaga, R., Umakanth, S., Bhat, D., & G S, S. K. (2024). Explainable artificial intelligence approaches for COVID-19 prognosis prediction using clinical markers. Scientific Reports, 14(1). https://doi.org/10.1038/S41598-024-52428-2

Data on COVID-19 (coronavirus) by Our World in Data. (n.d.). Retrieved February 24, 2024, from https://github.com/owid/covid-19-data/tree/master/public/data

Dipaola, F., Gatti, M., Giaj Levra, A., Menè, R., Shiffer, D., Faccincani, R., Raouf, Z., Secchi, A., Rovere Querini, P., Voza, A., Badalamenti, S., Solbiati, M., Costantino, G., Savevski, V., & Furlan, R. (2023). Multimodal deep learning for COVID-19 prognosis prediction in the emergency department: a bi-centric study. Scientific Reports 2023 13:1, 13(1), 1–10. https://doi.org/10.1038/s41598-023-37512-3

Dugas, A. F., Jalalpour, M., Gel, Y., Levin, S., Torcaso, F., Igusa, T., & Rothman, R. E. (2013). Influenza forecasting with Google Flu Trends. PloS One, 8(2). https://doi.org/10.1371/JOURNAL.PONE.0056176

Kingma, D. P., & Ba, J. L. (2014). Adam: A Method for Stochastic Optimization. 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings. https://arxiv.org/abs/1412.6980v9

Kucharski, A. J., Russell, T. W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., Eggo, R. M., Sun, F., Jit, M., Munday, J. D., Davies, N., Gimma, A., van Zandvoort, K., Gibbs, H., Hellewell, J., Jarvis, C. I., Clifford, S., Quilty, B. J., Bosse, N. I., … Flasche, S. (2020). Early dynamics of transmission and control of COVID-19: a mathematical modelling study. The Lancet. Infectious Diseases, 20(5), 553–558. https://doi.org/10.1016/S1473-3099(20)30144-4

Lorenzen, S. S., Nielsen, M., Jimenez-Solem, E., Petersen, T. S., Perner, A., Thorsen-Meyer, H. C., Igel, C., & Sillesen, M. (2021). Using machine learning for predicting intensive care unit resource use during the COVID-19 pandemic in Denmark. Scientific Reports 2021 11:1, 11(1), 1–10. https://doi.org/10.1038/s41598-021-98617-1

Loshchilov, I., & Hutter, F. (2016). SGDR: Stochastic Gradient Descent with Warm Restarts. 5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings. https://arxiv.org/abs/1608.03983v5

Loshchilov, I., & Hutter, F. (2017). Decoupled Weight Decay Regularization. 7th International Conference on Learning Representations, ICLR 2019. https://arxiv.org/abs/1711.05101v3

Oliva, R., & Oliva, © R. (1995). A Vensim® Module to Calculate Summary Statistics for Historical Fit D-4584 D-4584. http://metasd.com/wp-content/uploads/2010/03/D4584theil.pdf

Prioritizing diseases for research and development in emergency contexts. (n.d.). Retrieved March 14, 2024, from https://www.who.int/activities/prioritizing-diseases-for-research-and-development-in-emergency-contexts?gad_source=1

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving Language Understanding by Generative Pre-Training. https://gluebenchmark.com/leaderboard

Škraba, A., & Vavtar, B. (2022). Razvoj simulacijskega modela širjenja epidemije SI po principih agentnega modeliranja in sistemske dinamike. 945–956. https://doi.org/10.18690/UM.FOV.3.2022.69

Škraba, A., Vavtar, B., Stanovov, V., Semenkin, E., & Stojanović, R. (2021). Parametrization of bass diffusion model on COVID-19 first wave data. IOP Conference Series: Materials Science and Engineering, 1047(1), 012084. https://doi.org/10.1088/1757-899X/1047/1/012084

Stanovov, V., Grabljevec, S., Akhmedova, S., Semenkin, E., Stojanović, R., Rozman, Č., & Škraba, A. (2022). Identification of COVID-19 spread mechanisms based on first-wave data, simulation models, and evolutionary algorithms. PloS One, 17(12). https://doi.org/10.1371/JOURNAL.PONE.0279427

Sterman, J. (2000). Business Dynamics, System Thinking and Modeling for a Complex World. Http://Lst-Iiep.Iiep-Unesco.Org/Cgi-Bin/Wwwi32.Exe/[In=epidoc1.in]/?T2000=013598/(100), 19.

Stojanovic, R., Skraba, A., & Lutovac, B. (2020). A Headset Like Wearable Device to Track COVID-19 Symptoms. 2020 9th Mediterranean Conference on Embedded Computing, MECO 2020. https://doi.org/10.1109/MECO49872.2020.9134211

Subudhi, S., Verma, A., Patel, A. B., Hardin, C. C., Khandekar, M. J., Lee, H., McEvoy, D., Stylianopoulos, T., Munn, L. L., Dutta, S., & Jain, R. K. (2021). Comparing machine learning algorithms for predicting ICU admission and mortality in COVID-19. NPJ Digital Medicine, 4(1). https://doi.org/10.1038/S41746-021-00456-X

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention Is All You Need. Advances in Neural Information Processing Systems, 2017-December, 5999–6009. https://arxiv.org/abs/1706.03762v7

Wen, Q., Zhou, T., Zhang, C., Chen, W., Ma, Z., Yan, J., & Sun, L. (2022). Transformers in Time Series: A Survey. IJCAI International Joint Conference on Artificial Intelligence, 2023-August, 6778–6786. https://doi.org/10.24963/ijcai.2023/759

What is Disease X | Johns Hopkins | Bloomberg School of Public Health. (n.d.). Retrieved March 14, 2024, from https://publichealth.jhu.edu/2024/what-is-disease-x

Wright, S. (1921). Correlation and causation. Journal of Agricultural Research, 20(7), 557–585.

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2020). Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. 35th AAAI Conference on Artificial Intelligence, AAAI 2021, 12B, 11106–11115. https://doi.org/10.1609/aaai.v35i12.17325