Chapter xxx

How Weather Conditions Affect the Spread of Covid-19: Findings from a Study Using Contrastive Learning and NARMAX Models

Yiming Sun [1], **Hua-Liang Wei** [1,2 *]

[1] *Department of Automatic Control and System Engineering, University of Sheffield, Sheffield, S1 3JD, UK*

[2] *INSIGNEO Institute for in Silico Medicine, University of Sheffield, Sheffield, S1 3JD, UK*

* *Corresponding Author: w.hualiang@sheffield.ac.uk*

Abstract. Machine learning (ML) has demonstrated a powerful ability in learning complex patterns or inherent dynamics from observed data. Most machine learning models are black-box, in that the internal behaviour of the models is opaque and thus unknown to no one. However, in many real applications, e.g., in many medical and healthcare domains, it is significantly useful or necessary to explicitly know the internal compositions, combinations or interactions of the models to be used for one purpose or another. Therefore, the interest in interpreting machine learning models has increasingly grown in recent years, especially for cases where users need to do predictions using the models and require explanations for an insightful understanding of drivers that cause the predicted behaviour. This study introduces a novel interpretable machine learning method based on contrastive learning and Non-linear AutoRegressive Moving Average with eXogenous inputs (NARMAX) model (referred to as CL-NARMAX thereafter). The proposed method provides a glass-box model, where the input-output relationship and interactions between the input variables can be written down, so as the model cannot only be applied for predicting future behaviour but also for explaining the relevant “reasons” behind the predicted behaviour. Two case studies are provided to illustrate the usability and performance of the proposed CL-NARMAX approach. The first case study focuses on modelling and analyzing weather conditions against the Covid-19 data in the UK and France, aiming to reveal the impacts of climatic factors on the spread of Covid-19 using the proposed CL-NARMAX method. The second case study focuses on modelling the relationship between influenza-like illness (ILI) incidence rate and the relevant mortality based on the England data, where it is mainly served for illustration purpose, showing how CL-NARMAX is used to model a dynamic system, generating dynamic process models that can be used for explanation and prediction.

Keywords: NARMAX, Contrastive learning, COVID-19, Interpretable machine learning method.

1. Introduction

Machine learning (ML) has demonstrated its powerful ability in learning complex patterns or inherent dynamics from observed data, which has drawn considerable attention for its ability to predict complex phenomenon [1]. Most machine learning models are black-box, in that the internal behavior of the models is opaque and thus unknown to either the model builders or the end-users. However, in many applications there is a high demand for model’s accountability or explanability, that is, in addition to accurate prediction, ML should also be to tell, from the observed information and knowledge, the domain relationships contained in the data. This is referred to as interpretability [2]. In many application domains, for example, health care and medicine [3, 4], policymaking [5], and material design [6], researchers are concerned with not only the forecast of machine learning systems but also the explanation of machine learning models, or the relationship between the system outputs and inputs so as to make reliable decisions and provide clear guidance according to the explanation [7]. In healthcare and medicine, the interpretability of models and the interactions of system variables are usually treated as a prerequisite of using ML models, as the “reason” behind the prediction by ML models is usually most desirable and useful for making important decisions [8].

Coronavirus (Covid-19), the new global pandemic and the latest largest threat to global health, has been the focus of the past year [9]. There are totally over 192 million confirmed cases and over 4 million deaths according to the World Health Organization, while the new confirmed cases and deaths still keep rising up [10]. Thus, besides the prediction of new cases, it is vital for us to better understand the main factors that may influence the spread of the Covid-19 virus to prevent the spread of the virus [11]. As at the first stage of the pandemic, it was widely hoped that higher temperature would slow the spread of the virus in last summer [12]. But the reality shows that not only the temperate but also other factors may generate significant impact on the spread of the pandemic [13].

Since the start of the pandemic, a huge amount of research related to the prediction of new cases and spread have been carried out. A variety of methods have been proposed for predicting the number of new cases. These methods include ML, such as deep learning [14], support vector machine [15], fuzzy system [16], neural networks [17] and dynamical Bayesian models [18]. However, as mentioned earlier, most ML models can only generate a prediction of new cases without providing any information of the inner relationship of relevant factors.

This study introduces a novel interpretable machine learning method based on Contrastive Learning and Non-linear AutoRegressive Moving Average with eXogenous inputs (CL-NARMAX) model for medical data analysis. Unlike most ML methods which are black-box, the proposed method provides a glass-box model, where the input-output relationship and interactions between the input variables can be explained, so as the model cannot only be applied for predicting future new cases and deaths but also provide an explanation on how the spread of the spread of the pandemic is affected by different climatic and weather variables; such a model may be used for better understanding the pandemic dynamics and further studies in future.

The remainder of the paper is as follows. In Section2, the NARMAX model structured is described in detail. In Section3, the novel CL-NARMAX method is presented. In Section4, two case studies are provided, one focusing on modelling and analyzing weather conditions against the Covid-19 data in the UK and France and the other concerning the relationship between influenza-like illness (ILI) incidence rate and the relevant mortality The work is briefly summarized in Section5.

2. NARMAX Model

2.1 The Regression Model

For convenience of description, take NARX model (as a special case of NARMAX) as an example and consider a multivariate regression problem, with n predictor variables, $x_1, x_2, ..., x_n$, and one response variable y . The modelling task is to investigate the quantitative dependent relationship of the response on the predictors. Mathematically, the objective is to establish a model that links the predictors to the response via a function f as follows:

$$y(t) = f(x_1(t), x_2(t), ..., x_n(t)) + e(t) \tag{1}$$

where $x_i(\cdot)(i = 1, 2, ..., n)$ and $y(\cdot)$ represent the sequence of the observed predictor and response variables, respectively, $e(t)$ represents the model error; f represents some linear or non-linear functions. Usually, f is unknown but can be approximated from given observational data. There are a diversity of methods for building a function to approximate the true system, such as polynomials [19, 20], radial basis functions [21], and wavelet functions [22-25]. In this study, a polynomial based regression model is considered due to its superb properties [26]. By applying the polynomial form with the non-linear degree of up to ℓ , model (1) can be represented as:

$$
\begin{aligned}
y(t) = \theta_0 + \sum_{i_1=1}^{d} f_{i_1}(x_{i_1}(t)) + \sum_{i_1=1}^{d} \sum_{i_2=i_1+1}^{d} f_{i_1 i_2}(x_{i_1}(t), x_{i_2}(t)) \\
+ \sum_{i_1=1}^{d} \cdots \sum_{i_m=i_{m-1}+1}^{d} f_{i_1 i_2 \ldots i_m}(x_{i_1}(t), x_{i_2}(t), \ldots, x_{i_m}(t)) + e(t)
\end{aligned}
\tag{2}
$$

where $\theta_{i_1 i_2 \ldots i_l}$ are parameters, $d = n_y + n_u$ and

$$
f_{i_1 i_2 \ldots i_m}(x_{i_1}(t) x_{i_2}(t), \ldots, x_{i_m}(t)) = \theta_{i_1 i_2 \ldots i_m} \times \prod_{k=1}^{m} x_{i_k}(t), \quad 1 \le m \le \ell \tag{3}
$$

The degree of the multivariate polynomial is defined as the highest order among the terms. For example, the non-linear degree of the polynomial term $x_1(t)$ is 1, $x_1(t)x_2(t)$ is 2 and $x_1^2(t)x_2^3(t)$ is 5. Therefore, the degree of any term in model (2) is not higher than ℓ . In most practical implementations, the non-linear mapping function f can be approximated by a linear combination of a predefined set of functions $\phi_i(\varphi(k))$. Note that the polynomial NARX model described by Eq (2) can be written as the following linear-in-the-parameters form:

$$
y(t) = \sum_{m=1}^{M} \theta_m \varphi_m(t) + e(t) \tag{4}
$$

where $\varphi_m(t)$ are the model terms generated from the candidate variables, e.g., a model with the non-linear degree ℓ =2, involving only two variables, contains the following 6 terms: $\varphi_0(t) = 1$, $\varphi_2(t) = x_1(t)$, $\varphi_3(t) = x_2(t)$, $\varphi_4(t) x_1^2(t)$, $\varphi_5(t) = x_1(t) x_2(t)$, $\varphi_6(t) = x_2^2(t)$.

2.2 NARMAX Model Structure

Taking the case of a single-input and single-output system as an example, the NARMAX model is written as [25]:

$$\begin{aligned} y(t) = f\,(y(t-1),...,y(t-n_y),u(t-1),...,u(t-n_u), \\ e(t-1),...,e(t-n_e)) + e(t) \end{aligned} \tag{5}$$

where $u(t)$, $y(t)$,and $e(t)$ are the measured system input, output and noise signal respectively at time instant t ; n_u , n_y , and n_e are the maximum lags for the system input, output and noise; $f(\cdot)$ is some non-linear function to be identified. The NARMAX model (5) can be accommodated in the linear-in-the-parameters form (4) by defining $\varphi_m(t)$ as:

$$x_k(t) = \begin{cases} y(t-m) & 1 \le m \le n_y \\ u(t-m+n_y) & n_y + 1 \le m \le n_y + n_u \\ e(t-m+n_y+n_u) & \mathrm{n}_y + n_u + 1 \le m \le n \end{cases} \tag{6}$$

where $n = n_y + n_u + n_e$. Normally the noise signal $e(t)$ in model (5) is unmeasurable. Therefore, $e(t)$ is often replaced by the model residual sequence in model identification procedure:

$$e(t) = \varepsilon(t) = y(t) - \hat{y}(t) \tag{7}$$

where $\hat{y}(t)$ is the predicted value at time instant t generated by an estimated model. Detailed information about how to calculate model parameters and update model errors can be found in [27].

2.3 NARMAX Model Term Selection and Estimation

The initial linear-in-the-parameters model (4) may involve a large number of candidate model terms. However, only a small number of significant model terms are necessary in the final model to represent given observed data in most cases. Most candidate model terms are either redundant or make minimal contribution to the system output and can therefore be removed from the model. Thus, efficient model term selection and estimation method is needed.

The forward regression orthogonal least squares (FROLS) algorithms [20, 26] provides an efficient and powerful method for non-linear model term selection and model structure estimation. More detailed discussion of the FROLS algorithm and ERR index can be found in [28, 29]. In this paper, we only give a very brief summary of the algorithm. FROLS searches in a set consisting of all the specified possible candidate model terms or regressors to select the most significant model terms iteratively, in a stepwise manner, through an orthogonalization procedure, where the significance of model terms is measured by an index called the error reduction ratio (ERR) [26]. The FROLS algorithm will stop once the specified conditions are met, e.g., when the Error Signal Ratio (ESR) or index reaches its minimum [26]. Also, some statistical criteria, e.g., AIC [30], BIC [31], APRESS [32], can be used to monitor the model selection procedure and determine the model complexity.

3. Contrastive Learning Enhanced NARMAX Method

3.1 Contrastive Learning

Contrastive learning (CL), as a representation learning approach, has delivered impressive results on various scenarios [33]. This self-supervised training process can be understood as learning representation through contrastive positive data pairs against negative ones, where separate encoders could achieve the representation [34]. As the pre-definition or categorization for the candidate encoders is unnecessary, the process of CL is more effective and flexible.

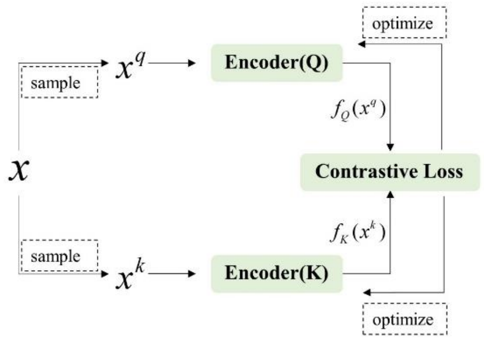

Figure 1 Contrastive Learning

The general training process of Contrastive Learning is shown in Figure 1. For general CL process, the input samples will be resampled into two independent parts: positive samples and negative samples by the related labels [35]. Normally, positive and negative samples are vital to the process of CL, while it is also possible to use the entire data set rather than to split it [36]. Note that the encoder Q generates representation for positive samples x^q and encoder K generates representation for negative samples x^k .When the contrastive loss between outputs of two encoders is converging, positive samples become closer, while negative samples get less similar.

Normally, contrastive learning focuses on comparing the embeddings with a Noise Contrastive Estimation (NCE) function that is defined as [33]:

$$L_{NCE} = -\log \frac{\exp(\cos\_sim(q,k_+)/\tau)}{\exp(\cos\_sim(q,k_+)/\tau) + \exp(\cos\_sim(q,k_-)/\tau)} \tag{8}$$

$$\cos\_sim(A,B) = \frac{A \cdot B}{\|A\|\|B\|} \tag{9}$$

where q is the original output, $k_+ = f_Q(x^q)$ represents the prediction by Encoder Q with the positive sample x^q , and $k_- = f_K(x^k)$ represents the prediction by Encoder K with the negative sample x^k , τ is a hyperparameter. More information of contrastive learning can be found in [37].

3.2 The CL-NARMAX Model

The scheme of the proposed framework is shown in Figure 2. Unlike the general CL process, the proposed CL-NARMAX modelling share the same data set $X_{D\times n}$ by two modelling frameworks, the true system (considering the true system as a model) and NARMAX models. The outputs of the two are models: $\boldsymbol{y} = F(\cdot)$ (observed system output) and $\boldsymbol{y}_l = F_{NARX}^l\,(l = 1,2,...,L)$ (model output), respectively. The system output $\boldsymbol{y}$ is set as the positive pair, and the outputs by NARMAX models

$\dot{\boldsymbol{y}}_l\,(l=1,2,\ldots,L)$ are set to be the negative pairs.

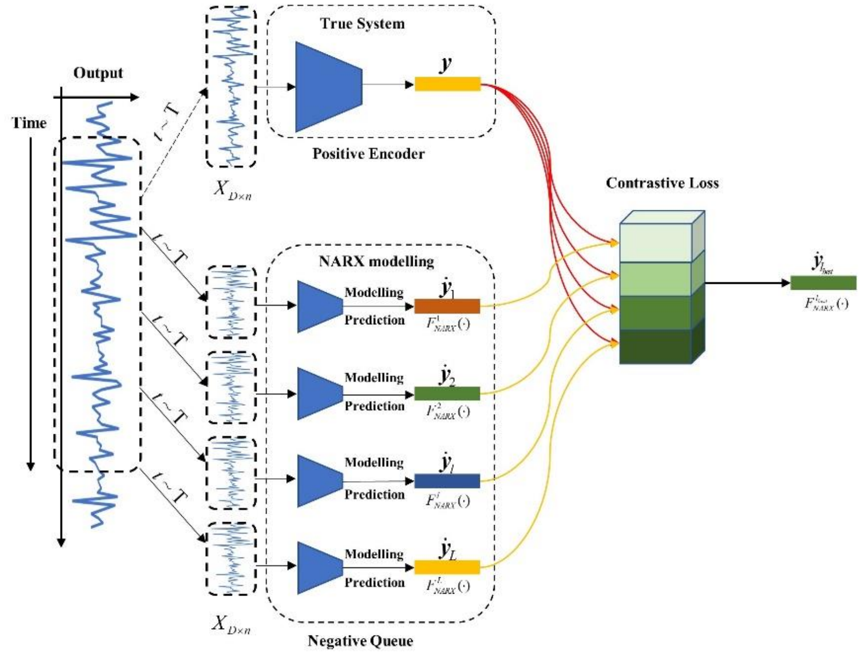

Figure 2 The General Process of CL-NARMAX

Notice that the identified models F_{NARX}^{l} and predicted values $\boldsymbol{y}_l$ by NARMAX vary due to the different choices of the parameters. Thus, the group generated by the NARMAX should be a set of possible options, which could also be called the negative queue. Then, for each negative pair in the negative queue, there will be a related contrastive loss with the positive pair, which is defined as follows:

$$CL_{NCE}^{l} = -\log \frac{\exp(h(\boldsymbol{y},\boldsymbol{y}))}{\sum_{l=1}^{L} \exp(h(\boldsymbol{y},\boldsymbol{y})) + \exp(h(\boldsymbol{y},\boldsymbol{y}_l))} \tag{10}$$

where $h(\boldsymbol{u},\boldsymbol{v}) = \exp(\boldsymbol{u}^T\boldsymbol{v} / \tau \cdot \|\boldsymbol{u}\| \cdot \|\boldsymbol{v}\|)$ is used to compute the similarity between $\boldsymbol{u}$ and $\boldsymbol{v}$ vectors with an adjustable parameter temperature, τ. Hence, the most similar negative pair compared with the positive pair is chosen as:

$$l_{best} = \arg\max_{1 \le l \le L} CL^{l}_{NCE} \tag{11}$$

where $L = n_y \times n_u \times \ell$, n_u and n_y are the maximum input and output lags of the NARMAX models; ℓ is the maximum non-linear degree of the NARMAX models. In this study here, the most effective NARMAX model, $F_{Nl_{best}}$ that generates the closest output $\tilde{y}_{\ell,best}$ to the corresponding measurement is included in the negative queue.

4. Case Studies

4.1 Modelling and Analyzing Weather Conditions Against Daily Confirmed Covid-19 Cases in the UK and France

4.1.1. Data

The daily confirmed cases for UK and France between 1st January 2020 and 27th July 2020 were acquired from the World Bank's World Development Indicators database [38], which is also known as the first wave of the pandemic. The dataset contains a total of 213 daily confirmed cases. Though there were no confirmed cases on 1st January in UK and France, that day was still assumed to be the starting time for the pandemic， and the raw data are plotted in Fig.1. Meanwhile, the daily meteorological data and the daily mobility data were collected from the DELVE program [38] and Apple mobility reports data respectively. The definition of the five climatic factors and two mobility variables, treated as input variables in this study, is summarized in Table 1.

A total of 213 data points for each country are split into two parts: the first 180 samples are used for model training and validation, while the rest 33 points are used for model testing. Note that each input variable does not impact the spread of the coronavirus immediately; previous study shows that their impacts become obvious after 7 days until 14 days [39]. The variation of the confirmed cases (denoted by y) is treated to be a dynamic process in this study. Therefore, a number of delayed output variables (autoregressive variables), namely, $y(t-1), y(t-2), ..., y(t-7)$, are also included in the model. Thus, to build an CL-NARMAX model, for inputs variables, the value of n_u varies from 7 to 14, while for output, n_y vary from 1 to 7. In addition, cross-product variables are also included in the model to reflect the impacts of the interactions between input and output variables on the spread of the virus. It argues that the impact of higher order interactions between input and output variables (e.g., higher than 3) on the pandemic becomes insignificant [40]. Thus, the degree ℓ of CL-NARMAX model is set to be 2.

Table 1 Climate and mobility variables

Variable	Symbol	Definition	Unit
Temperature	T	Average daily mean of temperatures	celsius Degrees
Humidity	H	Average daily humidity of air	Kilograms of water vapour per kilogram
Wind speed	Ws	Average daily wind speed	Meters per second
Solar radiation	Sr	Average daily short-wave radiation	W/m^2 (Watts per square meter)
Precipitation	P	Average daily precipitation	mm / hr
Driving	Dr	Percentage of change in routing requests by driving based on 100	N/A
Walking	W	Percentage of change in routing requests by walking based on 100	N/A

4.1.2. Results

The proposed CL-NARMAX model is applied to the training data of the two counties, and the identified best CL-NARMAX model for the UK case is:

$$\begin{aligned} y(t) = {} & 0.9104 y(t-1) + 0.0062 Dr(t-14) y(t-6) \\ & -96.0698 H(t-7) Sr(t-12) - 0.0187 Dr(t-8) \\ & +0.0126 Dr(t-6) + 0.0053 W(t-11) \end{aligned} \tag{12}$$

and the best CL-NARMAX model for the France case is:

$$\begin{aligned} y(t) = {} & 0.1365 P(t-9) + 0.0017 Sr(t-10) y(t-2) \\ & +0.0689 P(t-9) y(t-3) - 0.0089 Sr(t-13) \\ & +1.8559 y(t-6) \end{aligned} \tag{13}$$

Note that all the model terms involving noise variables such as $Sr(t-1)e(t-1)$ are omitted and not included in the final model, because all these noise terms are not useful for model prediction but are only used to reduce bias in model estimation. A comparison of the model predicted daily cases and the corresponding true values, on the training and test data sets, are shown in Figure 4 and Figure 5, respectively. Several statistical indicators are used for measuring the performance of the proposed method on training and test dataset, such as MSE (mean-squared-error), RMSE (root-mean-squared-error), MAE (mean absolute error), and R2 (coefficient of determination), where the values are listed in Table 2.

Model (12) (for the UK data) depends on the following four climatic factors: the humidity, solar radiation, driving and walking. More specifically, it shows that climate factors and mobility factors of one week ago have more influence on the present new cases. Similarly, model (13) shows that in France, the daily new cases is correlated to precipitation, solar radiation, and past cases, where only climate factors of one week ago make an impact on the present cases. From Figure 4, Figure 5 and Table 2, it can be seen that the proposed CL-NARMAX model shows an excellent prediction performance.

Table 2 Statistical indicators of proposed method for case 1

	MSE		RMSE		MAE		R2	
	UK	FR	UK	FR	UK	FR	UK	FR
Train	0.016	0.117	0.127	0.342	0.080	0.188	0.984	0.883
Test	0.100	0.569	0.998	0.754	0.677	0.407	0.974	0.414

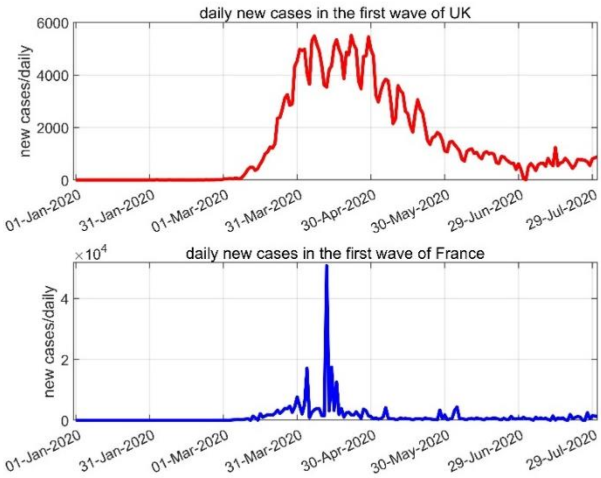

Figure 3 Daily confirmed cases in the first wave in UK and France

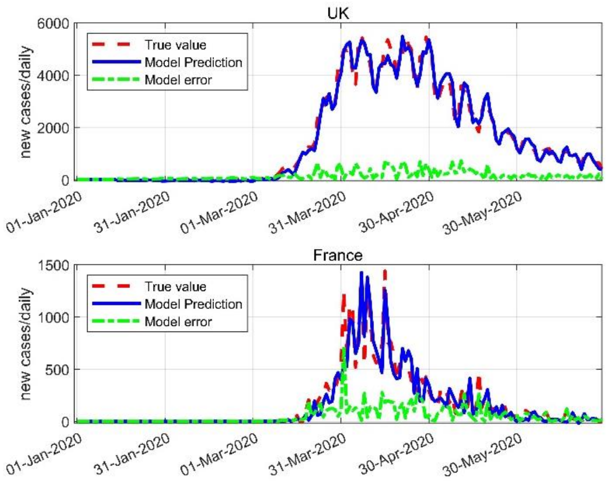

Figure 4. The comparison of the model prediction with the related true daily confirmed cases of UK and France on the training dataset.

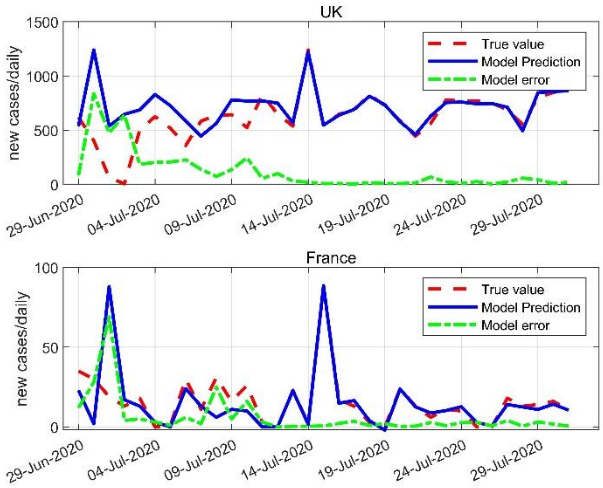

Figure 5. The comparison of the model prediction with the related true daily confirmed cases of UK and France on the test dataset.

4.2 The Relationship Between Influenza-like Illness (ILI) Incidence Rate and the Relevant Mortality

4.2.1. Data

The weekly influenza-like illness (ILI) incidence rate and deaths data of England were obtained from the Office for National Statistics (ONS), The Royal College of General Practitioners Research and Surveillance Centre and Public Health Wales. The dataset contains 835 weekly data points starting in week 31 of 2004 and ending in week 30 of 2020. The raw data are plotted in Figure 6.

Following the previous studies (e.g.,[41, 42]) The relationship between ILI incidence rate and weekly deaths is treated as a single-input single-output (SISO) system. Thus, in this problem, there is only one input variable, which is weekly ILI incidence rate. The number of weekly deaths is the system output. The initial dataset is split into two parts: the first 600 data points are for training and validation and the remaining 235 are for test.

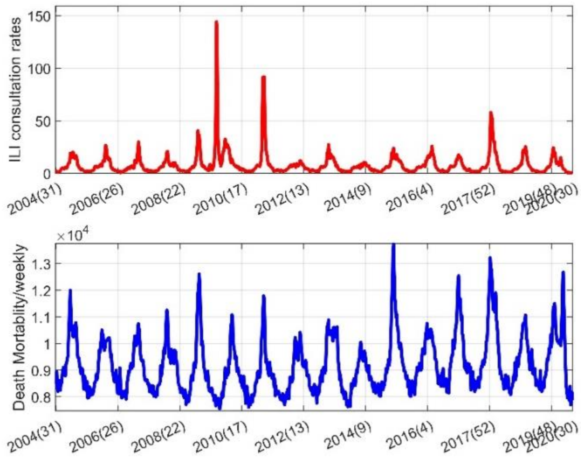

Figure 6 Weely ILI incidence rate and deaths of England from week 31 of 2004 to week 30 of 2020.

4.2.2. Results

The identified predictive model by the proposed CL-NARMAX method is as follows:

$$\begin{aligned} y(t) = & 1.0027y(t-1) - 0.0037u(t-3)y(t-1) \\ & +28.0887u(t-3) + 0.0005u(t-1)y(t-1) \end{aligned} \tag{14}$$

where $u(t)$ represents the weekly ILI incidence rate and $y(t)$ represents the number of weekly deaths. Similarly, there is no noise variables in model (14). A comparison between the model prediction and the corresponding observations on the training and test dataset are shown in Figure 7 and Figure 8. Similarly, several statistical indicators, such as MSE, RMSE, MAE, and R2, are used on training and test dataset, where the values are listed in Table 3.

From Figure 7, Figure 8 and Table 3, it can be seen that the identified CL-NARMAX model (14) shows an excellent prediction performance. More importantly, model (14) shows that the number of weekly deaths in the present week is closely related to the weekly ILI incidence rate one week and three weeks ago.

Table 3 Statistical indicators of proposed method for case 2

	MSE	RMSE	MAE	R2
Train	0.0936	0.3059	0.2318	0.9062
Test	0.0914	0.3023	0.2303	0.9082

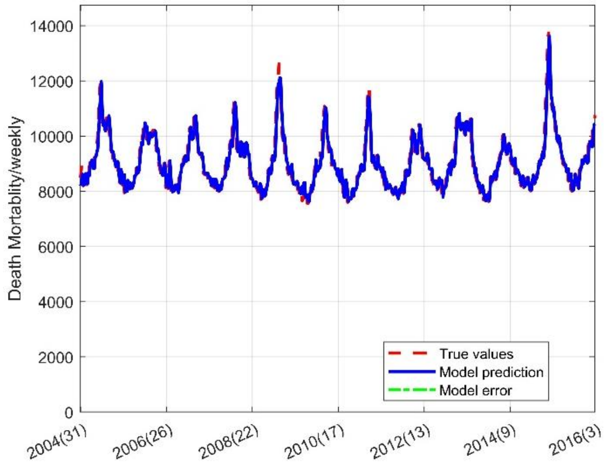

Figure 7 A comparison between the model prediction and the corresponding observations on the training dataset.

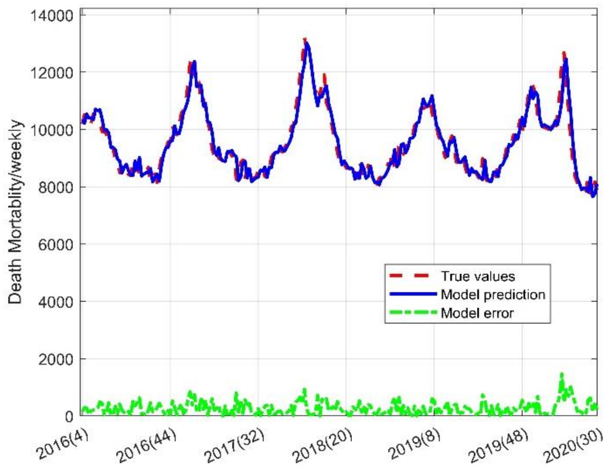

Figure 8 A comparison between the model prediction and the corresponding observations on the training dataset

5. Conclusion

This paper focuses on presenting a novel interpretable machine learning method based on contrastive learning and NARMAX model (CL-NARMAX) for solving health care data modelling problems. The proposed method can not only provide excellent predictions, but also possess the ability to explain how the output variable (response) is linked to the most important input variables (explanatory variables) and their interactions. The main contributions of the work are as follows. Firstly, the
proposed CL-NARMAX method takes advantage of contrastive learning and the NARMAX method, significantly improving the prediction ability of NARMAX models and meanwhile maintaining the distinct and attractive properties of NARMAX model e.g. the 'SIT' (sparse, interpretable and transparent) and 'SMART' (simple and simulatable, meaningful, accountable, reproducible, and transparent) properties [41, 42], which are significantly important and highly desirable in many real applications. Secondly, based on the data of the UK and France daily confirmation cases (both for the first wave), two model were provided, demonstrating how the climatic and weather conditions affected the spread of the coronavirus. These models can be useful for understanding the spread of the coronavirus.

This paper does not analyze model uncertainty and its effect on model generalization performance. Meanwhile, along with new variants and the development of vaccine since the start of 2nd wave of the pandemic, more possible factors affecting the spread of the virus shall be considered. Therefore, in the future, the model uncertainty of the proposed model will be further studied to improve the performance and robustness of this approach. Also, more candidate input variables will be investigated to develop more comprehensible model for the spread of the virus.

6. Acknowledgements

This work was supported in part by the Natural Environment Research Council (NERC) under the Grant NE/V001787 and Grant NE/V002511, the Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/I011056/1 and the Platform Grant EP/H00453X/1.

References

1. Vu, M.a.T., Adalı, T., Ba, D., Buzsáki, G., Carlson, D., Heller, K., Liston, C., Rudin, C., Sohal, V.S., Widge, A.S., Mayberg, H.S., Sapiro, G., and Dzirasa, K., A Shared Vision for Machine Learning in Neuroscience. J Neurosci, 2018. 38(7): p. 1601-1607.
2. Du, M., Liu, N., and Hu, X., Techniques for interpretable machine learning. Communications of the ACM, 2019. 63(1): p. 68-77.
3. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.a.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.a.W.M., Van Ginneken, B., and Sánchez, C.I., A survey on deep learning in medical image analysis. Med Image Anal, 2017. 42: p. 60-88.
4. Wu, H., Ruan, W., Wang, J., Zheng, D., Liu, B., Geng, Y., Chai, X., Chen, J., Li, K., Li, S., and Helal, S., Interpretable Machine Learning for COVID-19: An Empirical Study on Severity Prediction Task. IEEE Transactions on Artificial Intelligence, 2021: p. 1-1.
5. Brennan, T. and Oliver, W.L., The Emergence of Machine Learning Techniques in Criminology. Criminology & Public Policy, 2013. 12(3): p. 551-562.
6. Yadav, U., Pathrudkar, S., and Ghosh, S., Interpretable machine learning model for the deformation of multiwalled carbon nanotubes. Physical Review B, 2021. 103(3): p. 035407.
7. Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., and Pedreschi, D., A Survey of Methods for Explaining Black Box Models. ACM computing surveys, 2019. 51(5): p. 1-42.
8. Ahmad, M.A., Eckert, C., and Teredesai, A., Interpretable Machine Learning in Healthcare, in Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics. 2018, Association for Computing Machinery: Washington, DC, USA. p. 559–560.
9. Organization, W.H., Coronavirus disease (COVID-19). 2020.
10. Hassan, S.A., Sheikh, F.N., Jamal, S., Ezeh, J.K., and Akhtar, A., Coronavirus (COVID-19): A Review of Clinical Features, Diagnosis, and Treatment. Cureus, 2020. 12(3): p. e7355-e7355.
11. Mandal, M., Jana, S., Nandi, S.K., Khatua, A., Adak, S., and Kar, T.K., A model based study on the dynamics of COVID-19: Prediction

and control. Chaos Solitons Fractals, 2020. 136: p. 109889-109889.
12. Cruz-Cano, R., Ma, T., Yu, Y., Lee, M., and Liu, H., Forecasting COVID-19 Cases based on Social Distancing in Maryland, U.S.A.: A Time -Series Approach. Disaster Med Public Health Prep, 2021: p. 1-4.
13. Bukhari, Q., Massaro, J.M., D'agostino, S.R.B., and Khan, S., Effects of Weather on Coronavirus Pandemic. Int J Environ Res Public Health, 2020. 17(15): p. 5399.
14. 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, 2020. 12: p. 168-181.
15. Alimadadi, A., Aryal, S., Manandhar, I., Munroe, P.B., Joe, B., and Cheng, X., Artificial intelligence and machine learning to fight COVID-19. Physiol Genomics, 2020. 52(4): p. 200-202.
16. Pinter, G., Felde, I., Mosavi, A., Ghamisi, P., and Gloaguen, R., COVID-19 Pandemic Prediction for Hungary; A Hybrid Machine Learning Approach. Mathematics, 2020. 8(6): p. 890.
17. Babukarthik, R.G., Adiga, V.a.K., Sambasivam, G., Chandramohan, D., and Amudhavel, J., Prediction of COVID-19 Using Genetic Deep Learning Convolutional Neural Network (GDCNN). IEEE access, 2020. 8: p. 177647-177666.
18. Jha, P.K., Cao, L., and Oden, J.T., Bayesian-based predictions of COVID-19 evolution in Texas using multispecies mixture-theoretic continuum models. Computational mechanics, 2020. 66(5): p. 1055-1014.
19. Aguirre, L.A. and Billings, S.A., Dynamical effects of overparametrization in nonlinear models. Physica D: Nonlinear Phenomena, 1995. 80(1-2): p. 26-40.
20. Chen, S., Billings, S.A., and Luo, W., Orthogonal least squares methods and their application to non-linear system identification. International Journal of control, 1989. 50(5): p. 1873-1896.
21. Billings, S.A., Wei, H.-L., and Balikhin, M.A., Generalized multiscale radial basis function networks. Neural Netw, 2007. 20(10): p. 1081-1094.
22. Billings, S.A. and Wei, H.L., The wavelet-NARMAX representation: A hybrid model structure combining polynomial models with multiresolution wavelet decompositions. International journal of systems science, 2005. 36(3): p. 137-152.

23. Wei, H.L. and Billings, S.A., A unified wavelet-based modelling framework for non-linear system identification: the WANARX model structure. International journal of control, 2004. 77(4): p. 351-366.
24. Wei, H.L. and Billings, S.A., Long term prediction of non-linear time series using multiresolution wavelet models. International journal of control, 2006. 79(6): p. 569-580.
25. Wei, H.L., Billings, S.A., Yifan, Z., and Lingzhong, G., Lattice Dynamical Wavelet Neural Networks Implemented Using Particle Swarm Optimization for Spatio-Temporal System Identification. IEEE Trans Neural Netw, 2009. 20(1): p. 181-185.
26. Billings, S.A., Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains. 2013, Chichester, England: Chichester, England : Wiley, c2013.
27. Wei, H.L., Billings, S.A., Zhao, Y.F., and Guo, L.Z., An adaptive wavelet neural network for spatio-temporal system identification. Neural Netw, 2010. 23(10): p. 1286-1299.
28. Wei, H.L., Billings, S.A., and Liu, J., Term and variable selection for non-linear system identification. International Journal of Control, 2004. 77(1): p. 86-110.
29. Wei, H.L. and Billings, S., Model structure selection using an integrated forward orthogonal search algorithm assisted by squared correlation and mutual information. Int. J. Model. Identif. Control., 2008. 3: p. 341-356.
30. Akaike, H., Factor analysis and AIC, in Selected papers of hirotugu akaike. 1987, Springer. p. 371-386.
31. Burnham, K.P. and Anderson, D.R., Multimodel inference: understanding AIC and BIC in model selection. Sociological methods & research, 2004. 33(2): p. 261-304.
32. Billings, S.A. and Wei, H.-L., An adaptive orthogonal search algorithm for model subset selection and non-linear system identification. International Journal of Control, 2008. 81(5): p. 714-724.
33. Chen, T., Kornblith, S., Norouzi, M., and Hinton, G., A Simple Framework for Contrastive Learning of Visual Representations, in Proceedings of the 37th International Conference on Machine Learning, D. Hal, III and S. Aarti, Editors. 2020, PMLR: Proceedings of Machine Learning Research. p. 1597--1607.
34. Rao, H., Xu, S., Hu, X., Cheng, J., and Hu, B., Augmented Skeleton Based Contrastive Action Learning with Momentum LSTM for

Unsupervised Action Recognition. Information sciences, 2021. 569: p. 90-109.
35. Qin, S., Mudur, N., and Pehlevan, C., Contrastive Similarity Matching for Supervised Learning. Neural Computation, 2021. 33(5): p. 1300-1328.
36. Ashish, J., Ashwin Ramesh, B., Mohammad Zaki, Z., Debapriya, B., and Fillia, M., A Survey on Contrastive Self-Supervised Learning. Technologies (Basel), 2021. 9(1): p. 2.
37. Liu, Y., Li, Z., Pan, S., Gong, C., Zhou, C., and Karypis, G., Anomaly Detection on Attributed Networks via Contrastive Self-Supervised Learning. IEEE Trans Neural Netw Learn Syst, 2021. PP: p. 1-15.
38. Bhoopchand, A., Paleyes, A., Donkers, K., Tomasev, N., and Paquet, U., DELVE Global COVID-19 Dataset. 2020, The Royal Society.
39. Damette, O., Mathonnat, C., and Goutte, S., Meteorological factors against COVID-19 and the role of human mobility. PloS one, 2021. 16(6): p. e0252405-e0252405.
40. Lolli, S., Chen, Y.-C., Wang, S.-H., and Vivone, G., Impact of meteorological conditions and air pollution on COVID-19 pandemic transmission in Italy. Sci Rep, 2020. 10(1): p. 16213.
41. Wei, H.L., Sparse, Interpretable and Transparent Predictive Model Identification for Healthcare Data Analysis. 2019, Cham: Springer International Publishing: Cham. p. 103-114.
42. Billings, S. and Wei, H.L. NARMAX model as a sparse, interpretable and transparent machine learning approach for big medical and healthcare data analysis. in 2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS). 2019. IEEE.