Forecasting infectious and parasitic disease emergency department attendances using high-dimensional time series data

Peihong Guo
Lee Kong Chian School of Medicine
Nanyang Technological University
Singapore
peihong.guo@ntu.edu.sg

Pranav Tewari
Lee Kong Chian School of Medicine
Nanyang Technological University
Singapore
pranav.tewari@ntu.edu.sg

Esther Li Wen Choo
Lee Kong Chian School of Medicine
Nanyang Technological University
Singapore
L230001@e.ntu.edu.sg

Kelvin Bryan Tan
Policy Research and Evaluation Division
Ministry of Health
Singapore
kelvin_bryan_tan@moh.gov.sg

John Abisheganaden
Department of Respiratory and Critical Care Medicine
Tan Tock Seng Hospital
Singapore
john_abisheganaden@ttsh.com.sg

Borame Dickens*
Saw Swee Hock School of Public Health
National University of Singapore
Singapore
ephdbsl@nus.edu.sg

***Abstract*—The increasing global burden of infectious and parasitic diseases (IPDs) presents a significant challenge to public health systems worldwide, particularly to emergency departments (EDs). Accurate forecasting of ED attendances for IPDs is crucial for effective resource planning and public health interventions. We present a novel method for forecasting nationwide IPDs ED attendances incorporating a high-dimensional set of disease and environmental covariates. National ED attendances with IPDs as a cause in Singapore from 2009 to 2018 was taken as the study setting. We proposed simple and trimmed pooling of forecasts using 6 models which can optimize point forecast accuracy over a 1 to 8-week forecast horizon. We found that forecast combinations based on trimmed means provided superior forecast accuracy over sub-models which were used to construct the forecast combinations. Forecast combinations consistently achieved a Mean Absolute Percentage Error below 10.5% in the forecast horizon of 1–8 weeks ahead over the study period. We further demonstrated the robustness of forecast combinations in maintaining good forecast performance even when some sub-model constituents performed significantly worse compared to alternatives. Moreover, epidemiological analysis that employed post-selection inference revealed significant correlations between decreases in past temperature and increases in total precipitation, PM2.5 surface concentration with increased ED attendances for IPDs. The proposed methods can provide forward guidance for potential outbreaks and refine short to long–term resource management for ED admissions. The source code and supplementary material are available at https://github.com/gpeihong/IPDs-ED-Analysis.**

Keywords—forecasting, infectious diseases, forecast combinations, machine learning

I. Introduction

In recent years, the global burden of infectious and parasitic diseases has increased significantly, posing a tremendous challenge to public health systems worldwide [1]. These diseases, which include influenza, malaria, dengue, and various other viral, bacterial, and parasitic infections, have led to millions of deaths and a substantial impact on economies and healthcare systems [2]. The urgency of managing and controlling infectious and parasitic diseases is further exacerbated by the ongoing threat of pandemics and emerging pathogens [3].

One critical aspect of the public health response to infectious and parasitic diseases is the management of emergency care demand. Emergency departments (EDs) are often viewed as the primary entry point to a hospital, as they serve as an essential stop for patients before admission to various hospital services [4]. EDs play a vital role in providing timely and effective treatment to patients, as well as in identifying and containing potential outbreaks by isolating individuals infected with emerging diseases. Across the years, there has been a consistent increase in the demand for ED services. As a result, EDs face considerable pressure due to the high patient flow, making them some of the most congested areas within hospitals [5]. Managing the demand for ED services is a challenge encountered by numerous hospital systems. Various studies on EDs have shown that these facilities increasingly struggle to fulfill their responsibilities. Poor management of overcrowding can lead to inadequate ED functioning [6], potentially resulting in negative patient outcomes [7].

As the demand for emergency medical care continues to grow, it is crucial for EDs to integrate an adequate degree of resilience by incorporating forward guidance into their operations. This will allow them to foresee ED care requests and quickly mobilize essential resources [8]. Throughout the years, numerous methods have been developed to enhance modeling and forecasting of emergency department ED demands. Time series models are the most prevalent among the various models employed for forecasting ED demands [9]. These methodologies encompass autoregressive integrated moving average models (ARIMA) and its derivatives, as well as Holt-Winters methods [10]. The majority of existing research employing time series models utilized multivariate ARIMA models for single-step forecasting or univariate ARIMA models for multi-step forecasting. [10] presented a method using a multivariate ARIMA model to forecast ED visits at Lille hospital in France. [11] suggested a univariate

*Corresponding Author
This research is supported by the LKC School of Medicine, NTU, under the Staff Training and Development.

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00119

multi-step forecasting approach for electricity load based on the ARIMA model. A primary disadvantage of time series models is that time series models and their extensions may not achieve satisfactory performance when the time series data hardly exhibit regular variations. To counter this drawback, more adaptable non-parametric models and machine learning techniques can be employed to enhance forecasts. These comprise of and are not limited to statistical models using meteorological variables, including generalized additive models, etc. [12], traditional machine learning models such as the Random Forest (RF) regressor and deep learning methods such as the recurrent neural network (RNN). For traditional machine learning methods, length of stay for inpatients in a hospital has been predicted using Naive Bayesian classifiers [13]. Random forest regression has also been applied to forecast the arrival of emergency department patients by incorporating both meteorological and calendar information [14]. For deep learning methods, [15] carried out a comparative analysis of various deep learning techniques to forecast ED visits and the results highlighted the promising performance of these deep learning models, with the Variational AutoEncoder (VAE) outperforming the rest. Several deep learning-based combination approaches, including a combination of Gated Recurrent Units and Convolutional Neural Networks, Generative Adversarial Networks-based Recurrent Neural Networks and Convolutional LSTM Network, have also been developed to leverage the benefits of different models to enhance forecast accuracy [16, 17]. While these studies have achieved satisfactory predictive results, they have neglected the combination of traditional machine learning models. Traditional machine learning methods can not only perform well on relatively small datasets, but they also provide insights into the relationships between input features and target variables, making them generally easier to interpret and explain. Therefore, forecast combinations that combine baseline time series models and ensemble machine learning approaches [18] and combine large sets of environmental variables can potentially enhance forecast accuracy of disease burden [19].

In this paper, we included meteorological and atmospheric covariates as well as time series of other diseases' ED attendences for forecasting nationwide weekly infectious and parasitic diseases (IPDs) emergency department attendances. In addition, to guide short– and long–term ED demand management, multi–step forecasts were considered in this study using 6 sub-models, including baseline naïve forecasts, time series linear models, ensemble machine learning models, and deep neural networks. Moreover, among the submodels, we also extended the classical deep learning models LSTM and CNN for direct ED attendance point forecasts. We further explored the use of forecast combinations based on simple averaging and trimmed means. All submodels and forecast combinations specifically optimized forecasting accuracy over a 2-month time horizon. We evaluated forecast combinations against sub-models constituents on point forecast accuracy in the 1–8 week ahead forecast horizons. We demonstrated that our forecasting framework was able to generate predictions quickly and accurately, and that its deterioration rate did not increase significantly as the forecast horizon increased from 1 to 8 weeks.

II. Materials and Methods

A. Disease Surveillance Data

Nationwide daily emergency department attendances for a total of 25 Singapore burden of disease categories such as infectious and parasitic diseases, respiratory infection (RI), malignant neoplasms, skin diseases and diabetes mellitus were obtained from the Ministry of Health, Singapore from January 1, 2009, to December 31, 2018. Although most respiratory infections (RIs) are caused by bacteria, viruses, and fungi, some less common cases result from parasites. One example is paragonimiasis, a lung fluke infection caused by paragonimus species. Infections with these parasites can lead to symptoms such as cough, chest pain, and difficulty breathing [20]. We included all 25 categories in the set of predictive covariates and aggregated the temporal resolution of the data to epidemiological weeks (EW) comprising EW1 2009 to EW52 2018.

B. Meteorological and Ambient Air Pollution Data

We collected a total of 25 well-established environmental variables known to influence virus and vector survival, as well as parasite fecundity. These variables included daily total rainfall, mean, maximum, and minimum temperature, wind speed, vapor pressure, etc. Meteorological data were obtained from ERA5-Land, published by the European Centre for Medium-Range Weather Forecasts. Data related to ambient air pollution was obtained from National Environment Agency, which compiles air quality details for Singapore from 2009 to 2018, on a daily basis. The data included measurements for 5 major ambient air pollutants: $PM_{2.5}$, NO_2, O_3, SO_2, and CO. Subsequently, meteorological and ambient air pollutant exposures values were aggregated at a weekly level and their averages were calculated.

C. Point Forecast Submodels and Forecast Combinations

We aim to generate point forecasts for IPDs ED attendances 1–8 weeks ahead. The high-dimensional set of covariates here include 4 weeks lag of ED attendances for all 25 diseases burden categories, meteorological variables as well as air pollutant variables collected in the data sources described above (Refer to supplementary material). A lag-order of four was chosen for all variables, to encompass the potentially large generation interval of disease transmission, and further lag orders were not included for model parsimony.

We considered an autoregressive linear model (AR), using only past lags of IPDs ED attendances as predictors. Least-squares-based methods may not only tend to overfit but also lead to high prediction variance when a large number of covariates are incorporated. Therefore, in addition to using AR, we also proposed multivariate linear methods that incorporate regularization, ensemble-based methods, and deep neural networks as described below.

We augment the AR model by including exogenous variables (AREV). We regard the following conditional expectation E(.|.) as the h-week ahead forecast for IPDs ED attendances.

$$\begin{aligned} E(y_{t+h,D^*}|X,\beta) = \beta_0 &+ \sum_{l=L_{D^*}\in\{0,1,2,\dots\}} \beta_{l,y} y_{t-l,D^*} \\ &+ \sum_{d\in D} \sum_{l=L_D\in\{0,1,2,\dots\}} \beta_{l,d,y} y_{t-l,d} \\ &+ \sum_{m\in M} \sum_{l=L_M\in\{0,1,2,\dots\}} \beta_{l,m,y} x_{t-l,m}. \end{aligned} \quad (1)$$

Where t represents the current time point, and y_{t+h,D^*} denotes the weekly IPDs ED attendances at $t+h$. It is important to note that we use D to represent the set of diseases considered as predictors, excluding IPDs, and D^* as the index for weekly IPDs ED attendances. Autoregressive terms incorporate past and current observations of IPD ED weekly attendances, denoted as y_{t-l,D^*}, where l represents the lag order up to a maximum of L_D lags. The transmission of a disease of interest may also be partially explained by factors related to other diseases. Consequently, additional predictors of IPDs include ED attendances of other infectious diseases, represented as $y_{t-l,d}, d \in D$. In addition, the influence of environmental factors on the future burden of disease is also significant. The environmental covariates at t-l are denoted as $x_{t-l,m}$, where m represents the meteorological and air pollutant variables. $X=\{y_{t-l,D^*}, y_{t-l,d}, x_{t-l,m}\}$ represents the predictor matrix and $\beta \in \{\beta_0, \beta_{l,y}, \beta_{l,d,y}, \beta_{l,m,y}\}$ denotes the set of coefficients. We assumed that y_{t+h,D^*} follows a normal distribution, and the conditional mean of y_{t+h,D^*} serves as the h-week ahead point forecast.

We considered the Least Absolute Shrinkage and Selection Operator (LASSO) strategy to estimate the coefficients β and the set of parameters $\{D^*, D, M\}$ of AREV specification. LASSO induces variable shrinkage and sparsity by simultaneously selecting which parameters to include in the model and determining their appropriate values. The objective function of LASSO is as follows:

$$\underset{\beta}{\operatorname{argmin}} \ (y - X\beta)^2 + \lambda\|\beta\|, \qquad (2)$$

where it represents the total squared difference between the dependent variable y and the multiple between predictor matrix X and coefficients β, as well as a penalty term controlled by a parameter λ. Larger λ induces greater variable shrinkage and sparsity. The optimal λ is selected at the value which minimizes prediction error of the model in the training set through cross-validation. Therefore, models calibrated with the optimal regularization parameters tend to provide more robust predictions.

In addition to regularization, we also explored the Random Forest (RF) method. RF was originally developed by [21] and can be applied to either classification or regression. The RF method for regression is a predictor created by aggregating the predictions of many regression trees. The bootstrap bagging method is employed to select random sample sets from the training data set, fit multiple regression trees to these samples, and then aggregate the predictions from each tree. The bootstrap method improves model performance by reducing the prediction variance of the model without increasing bias or error. After training, the Random Forest prediction for a test sample is obtained by taking the average of the predicted values from all individual trees. As this study required 6 sub-models to respectively forecast ED attendance 1–8 weeks ahead, and we used a cross-validation strategy for each model's forecasting, which consumed substantial computational resources and time costs, we used the default parameter settings for all sub-models.

We also explored using gradient boosting machines (GBM). GBM is an ensemble learning model that uses decision trees sequentially, with each tree trained on the bias of previous trees. It aggregates the set of predictions from each tree to form an aggregate forecasts. In GBM, new decision trees are created by reducing existing forecast residuals. This process was carried out iteratively until the discrepancy between the predictions and the actual data surpasses a predetermined threshold.

We also employed Long short–term memory networks (LSTM) combined with convolutional neural networks (CNN) to forecast IPD ED attendances. LSTM can retain prior network information and connect it with current data, resulting in superior performance, while CNN networks are recognized for their capacity to automatically identify important features, enabling them to learn the internal structures of time-series data [22]. We therefore considered combining the features of these two deep learning techniques. A hybrid deep learning architecture called the LSTM-CNN model (Fig. 1) was constructed with LSTM and CNN combined and trained concurrently. The CNN then acts on the LSTM's output, further extracting and compressing these temporal features into a spatial feature map.

Increasing the number of LSTM layers and hidden units can allow the network to learn more complex patterns in the input data, but it also increases the computational cost of training the network. However, as the dataset in this study is small, we designed only 2 layers for the LSTM module [23]. The rationale for the convolution module to also be designed with 2 convolutional layers is to extract deep features [24], as these layers are highly resistant to noise. One-dimensional convolutional layers were employed, which effectively extract location-invariant features from short segments of the time series. During the training process, we divided the training sequence of each cross-validation fold into a training set and a validation set in the ratio of 6:1 respectively, where the validation set can help us select the appropriate parameters during the training process. Mean Squared Error (MSE) was used as the loss function, and the Adam algorithm [25] served

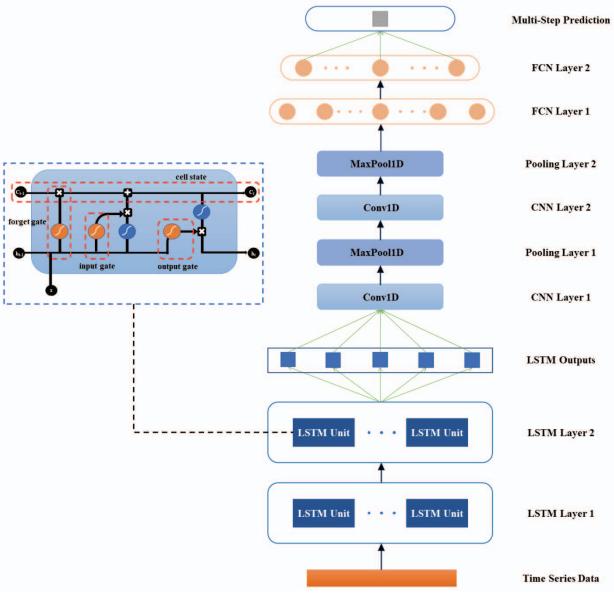

Fig. 1. **Hybrid LSTM-CNN structure.** The LSTM module consists of 2 layers with 16 hidden units each. The CNN includes 2 convolutional layers with kernel sizes of 3 and 2, both using a stride of 1 and Rectified Linear Unit (ReLU) as the activation function. The dashed box depicts the internal structure of the LSTM unit: x_t, c_t, and h_t represent the input information, cell state, and hidden information at t, respectively. The orange and blue circles correspond to the sigmoid and tanh functions, respectively.

as the optimizer. Within the Adam algorithm, L2 regularization is incorporated to prevent training overfitting, with a regularization parameter of 0.001.

We also considered a naive forecast, using the latest observations of IPDs ED attendances as the point forecast. Naive forecast is also often used as a baseline to compare with forecasts generated by more sophisticated methods. It tends to work well in situations where trends in the data remain relatively constant over time [26].

Lastly, we proposed two forecast combinations based on simple averaging (CombM) and trimmed mean (CombT), respectively. The trimmed mean is used to remove extreme and outlying predictions. Here, we set the trimming percentage to 20% to remove the outlying maximum and minimum value of forecasts generated from the 6 submodels and calculated the average of the remaining data. Forecast combinations combine the forecasts from Naive, AR, AREV (LASSO), LSTM-CNN, RF, GBM. This approach can integrate the predictive information of each model, potentially resulting in more robust and consistent forecast performance.

D. Optimization of Forecast Accuracy

To incorporate the latest information into model estimation, we performed cross-validation for all sub-models. An extended window splitting method was employed to evaluate forecasts. First, we divided the full dataset into an initial training set and a full forecast set to comprise 70% and 30% of the data respectively. Following this, folds over time were generated using an expanding window, where the length of the training sequence was increased by 1 week each time. Also, the forecast set is the subsequent epidemiological week of the current training sequence, whose length is always 1. At each time point, individual sub-models will be calibrated for each forecast time step, for each forecasting framework, to generate 1 to 8 weeks-ahead direct forecasts. This procedure was done until the final week of the dataset is incorporated for assessment. The accuracy of each model's forecast was evaluated by comparing the discrepancy between the forecasted and observed values in the full forecast set.

E. Performance Evaluation of Models

We used the mean absolute percentage error (MAPE), mean absolute forecast error (MAFE), and mean absolute scaled error (MASE) to evaluate the accuracy of ED attendances forecasting. The summary statistic MAPE summarises the percentage error each forecasting model makes relative to the actual observations. MAFE measures the absolute difference between forecasted and actual values, while MASE compares a model's accuracy to a naive forecast. A value less than 1 implies better performance than the naïve forecast, while a value greater than 1 suggests worse performance. Furthermore, the Diebold-Mariano hypothesis test was performed pairwise among models to statistically determine whether the forecasts were equivalent or non-equivalent throughout all forecast horizons at a 5% significance level [27].

III. Results

A. Summary Statistics in Study Setting

Fig. 2 shows the trend of some of the variables over time in our study setting. From 2009 to 2018, the case counts of ED attendances for IPDs in Singapore were relatively stable in the range of 900-2000 (Fig. 2A). The case counts of RIs ED attendances also fluctuated without displaying any discernible

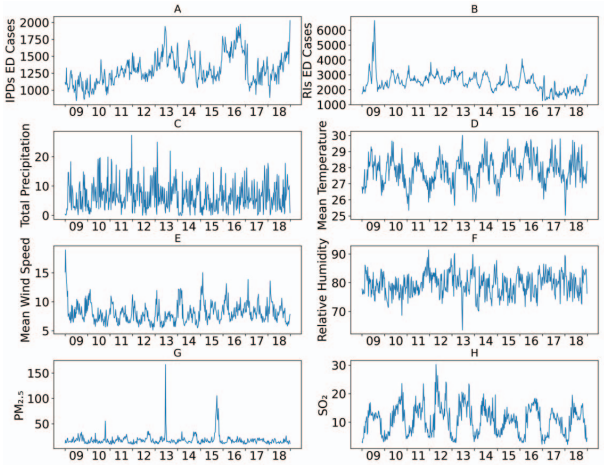

Fig. 2. **Weekly time series data from epiweek 1 2009 to epiweek 52 2018** on (A, B) nationwide emergency department attendances for infectious and parasitic diseases (IPDs) and respiratory infections (RIs), (C) average daily total precipitation (Millimeter), (D, E, F) average daily mean temperature (Celsius), wind speed (Kilometer/hour), relative humidity (%) and (G, H) average daily PM2.5 and SO2 surface concentration (micrograms per cubic meter).

pattern. However, in 2009, a significant increase in RIs ED attendances was noticed and may be attributed to the H1N1/09 pandemic. Afterward, the ED attendances gradually returned to pre-pandemic levels by 2010 (Fig. 2B).

Meteorological variables remained relatively constant overall with an average value of 6.520mm (Fig. 2C, Range: 0.000mm – 27.362mm), 27.840° C (Fig. 2D, Range: 25.040 ° C – 30.015° C), 8.130km/h (Fig. 2E, Range: 5.145km/h – 18.954km/h) and 79.460% (Fig. 2F, Range: 63.546% – 91.475%) for average daily total precipitation, average daily mean temperature, mean wind speed and relative humidity respectively. $PM_{2.5}$ (Fig. 2G) had a relatively high peak in the middle of the years 2010, 2013, and 2015, but was relatively constant otherwise. As for the SO_2, aside from a peak in the middle of 2012, concentrations remained relatively constant (Fig. 2H).

B. Overall Performance of Forecast Models

For each forecast method, 6 sub-models were trained using covariates with 4 weeks lags as predictors, with each model respectively generating 1 to 8-step direct forecasts. In general, forecast error increased as the forecast time horizon expanded (Fig. 3). However, different forecast models had noticeably different rates of performance deterioration. We found that AR, AREV (LASSO) and forecast combinations were the best performing models in the 1–3 weeks horizon (Fig. 3A–3C). After 3 weeks, the forecast combinations and GBM showed superior forecasting performance in comparison to the other models. While RF did not perform well in the 1–3 weeks horizon, it had similar forecasting performance as the forecast combinations in the 4-week or more forecast horizon (Fig. 3A–3C). AR and AREV (LASSO) performed relatively poorly in the forecast horizon of 4 to 8 weeks ahead, but mean absolute percentage errors for both models consistently stood below 11.5% (Fig. 3A).

At 4-week, the MAPE of RF decreased (Fig 3A: MAPE at 3 and 4 weeks: 9.10%, 9.00%), and at 7-week, the same pattern was observed for GBM (Fig 3A: MAPE at 6 and 7 weeks: 9.25%, 8.94%), which highlights that the RF may be

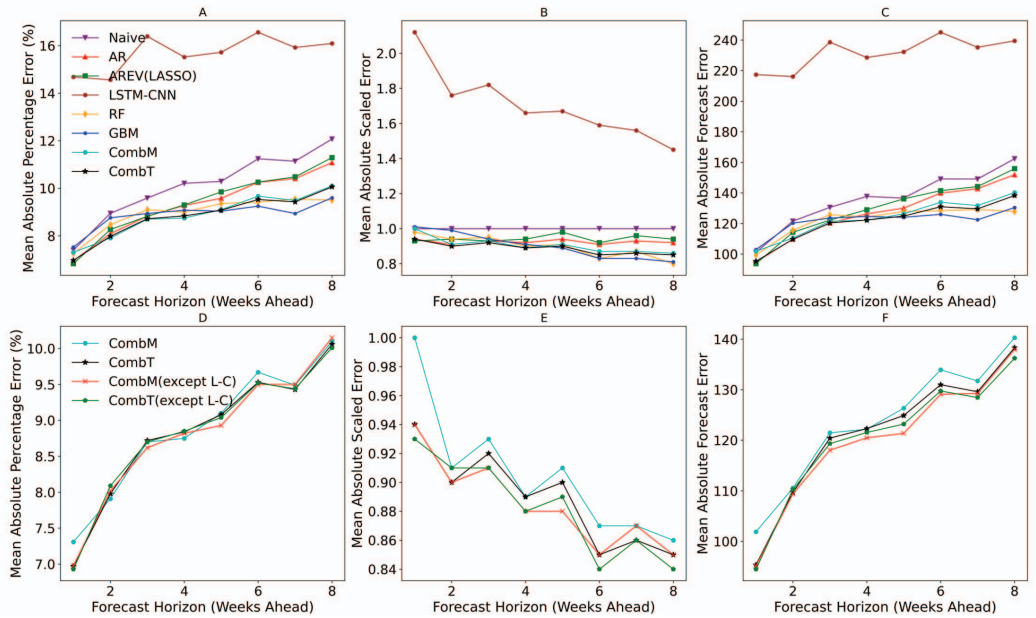

Fig. 3. **Evaluation of forecasts throughout the full forecast dataset.** First row: comparing forecast performances of the 8 forecasting models, which include the baseline Naive forecast, baseline AutoRegressive (AR) model, AutoRegression with Exogenous Variables (AREV) using LASSO, LSTM-CNN, random forest (RF), gradient boosted machines (GBM), a simple mean of all forecasts (CombM) and a trimmed mean of all forecasts (CombT) across the forecast horizons of 1 – 8 weeks ahead using (A) mean absolute percentage forecast error, (B) mean absolute scaled error and (C) mean absolute forecast error. Second row: comparing the forecast performance of forecast combinations that include LSTM-CNN with the forecast combinations that exclude LSTM-CNN (L-C) across the forecast horizons of 1 – 8 weeks ahead using (D) mean absolute percentage forecast error, (E) mean absolute scaled error and (F) mean absolute forecast error.

more suitable for medium and long–term forecasting and GBM may be more suitable for long–term forecasting (Fig. 3A, 3C). In addition, the mean absolute scaled errors of all models except LSTM-CNN were below 1 in all considered forecast horizons, demonstrating outperformance over the naive forecast (Fig. 3B). For LSTM-CNN, forecast performance was noticeably worse compared to all other models in all considered horizons. This may be due to LSTM-CNN requiring more data to adequately calibrate the model, which cannot be fulfilled given the small sample size of training data.

Given the poor forecasting performance shown by LSTM-CNN, it was necessary to explore whether its use as a sub-model constituent affects the performance of forecast

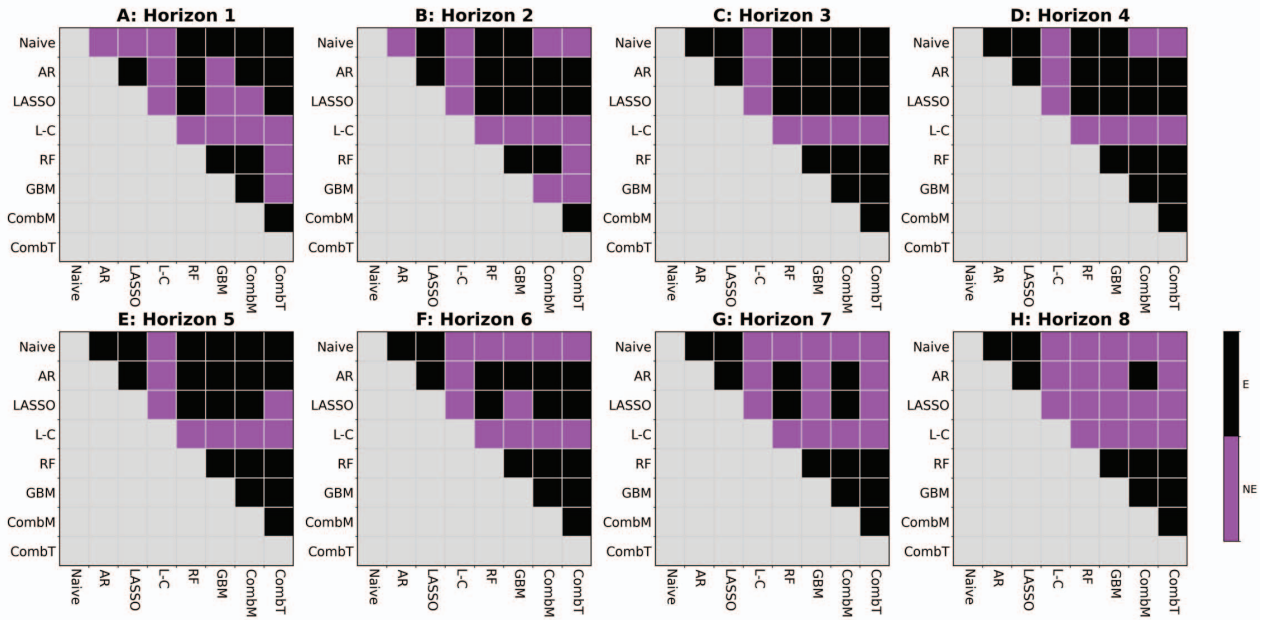

Fig. 4. **Visual representation of the Diebold-Mariano (DM) test for assessing the statistical equivalence of forecast errors across different models and forecast horizons.** Distinct panels indicate forecast non-equivalence (NE) in purple or equivalence (E) in black for a particular forecasting horizon. This was computed for forecast residuals in the full forecast set in the Naive forecast, AutoRegressive (AR) model, AutoRegression with Exogenous Variables using LASSO (LASSO), LSTM-CNN (L-C), random forest (RF), gradient boosted machines (GBM), a simple mean of all forecasts (CombM) and a trimmed mean of all forecasts (CombT).

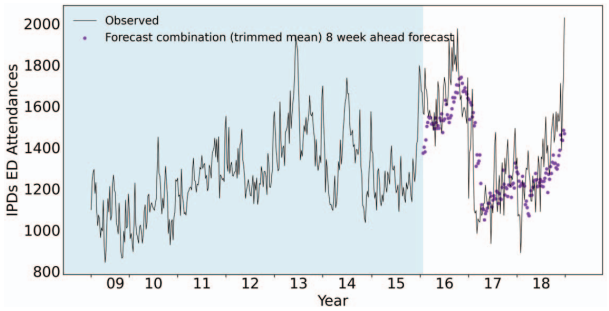

Fig. 5. Visualization of IPDs ED attendance case counts versus 8-week ahead forecasts from forecast combinations based on trimmed mean (CombT).

combinations. We compared the forecast error of the two forecast combinations schemes before and after excluding LSTM-CNN. We found that error rates for forecast combinations in both cases were not significantly different. In addition, we also found that the forecast accuracy of forecast combinations using the trimmed mean approach is consistently slightly higher than that of the simple averaging approach in all forecast horizons (Fig. 3D–3F), which highlights that trimming extreme forecasts can help improve forecast performance.

We employed the Diebold-Mariano (DM) test to examine whether forecast errors were equivalent across forecast models and forecast combinations. The hypothesis tests indicated that there were significant differences in forecast errors among models at a 5% significance level throughout all forecast horizons. In particular, the GBM model's performance was significantly superior to at least half of the forecasts produced by other models for horizons of 7 to 8 weeks (Fig. 4G–4H; Fig. 3A–3C). This finding also indirectly suggested that the GBM model may have been more appropriate for long–term forecasting. AR and AREV (LASSO) were always equivalent in the horizon of 1–8 weeks. (Fig. 4A–4H). Forecast combinations based on trimmed means consistently performed significantly better than most forecast models, especially in the 1–3 weeks and 7–8 weeks horizon (Fig. 4A–4B; Fig. 4G–4H; Fig. 3A–3C).

C. *Forecast Performance and Errors by Time*

Since forecast combinations integrated the strengths of various forecast sub-models and yielded smaller forecast errors, we focused on examining forecasts generated by forecast combinations which aggregated forecasts using the trimmed mean across time (See supplementary material for forecasts from each forecasting model and each horizon).

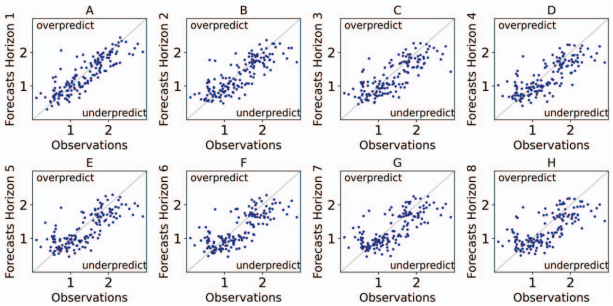

Fig. 6. **Visual representation of the forecasts produced through forecast combinations based on trimmed mean compared to the observed data in the full forecast dataset.** Each panel illustrates the comparison between forecasts and observations for each forecast horizon within 1 – 8 weeks.

Throughout 2016, a substantial discrepancy between forecasts and observations was observed for 8-week ahead forecasts (Fig. 5).

The majority of forecast errors were evenly dispersed along the equality line throughout the time series, suggesting that the forecasts did not exhibit significant bias versus observations (Fig. 6A–6H). However, forecasts in 2016 were consistently underpredicted and led to large, negative forecast errors in this period. Overall, underprediction was most notable for the forecast horizon of 3 to 8 weeks ahead (Fig. 6C–6H).

D. *Effects of Environmental Variables on Forecasts*

To explore the effects of environmental variables on IPDs ED attendances, we used LASSO to first select predictors from the set of covariates used for forecasting. Thereafter, a linear model was fitted to the LASSO-selected subset and confidence intervals for each regression coefficient were obtained by post-selection inference. As it is difficult to conduct inference on the parameters in a high-dimensional linear model, we only described general patterns of associations between environmental variables and IPDs ED attendance for all horizons (See supplementary material).

Increases in past total precipitation increased IPDs ED attendances, but the regression coefficients were only significant at the 5% level for the 1-week ahead of the forecasting window. A 1 millimeter increase in the past 1-week total precipitation is associated to a 3.86 (95% CI: 1.97 – 5.75) expected increase in IPD ED attendance. In contrast, increases in past $PM_{2.5}$ were associated with increased IPDs ED attendances for the 1- and 6-week ahead forecasting window. For instance, in the 6-week ahead horizon, a 1 microgram per cubic meter increase in the past 4-week PM2.5 surface concentration is associated to a 1.30 (95% CI: 0.03 – 2.57) expected increase. Increases in CO (coefficient: -203.07, 95% CI: -374.10 – 32.05) and O_3 (coefficient: -2.38, 95% CI: -4.32 – -0.43) were associated with decreases in IPDs ED attendances only in the 2-week forecasting window. Increases in past temperature and wind speed were associated with decreased IPDs ED attendances. For past temperature, the relationship held in the 1–5 weeks window. For example, in the 1-week ahead horizon, a 1 degree increase in the past 2-week minimum temperature is associated to a 31.90 (95% CI: -46.79 – -17.02) expected decrease. However, the relationship between past wind speed and IPDs ED burden only held in the horizon of 8-week ahead. A 1 kilometer per hour increase in the past 4-week max wind speed is associated to a 6.75 (95% CI: -12.08 – 1.41) expected decrease.

IV. Discussion

Our results demonstrated that our proposed trimmed mean-based forecast combinations yielded the best performance in the forecast horizon of 1–8 weeks ahead for the full forecast dataset. Forecast errors under this forecast combination scheme were less than 10.5% in the 1 – 8 week ahead forecast horizons. The proposed forecasting framework can therefore offer substantial value in supporting resource management in EDs, by providing forward guidance on the estimated number of ED attendances related to IPDs. This can help aid proactive management of crowding situations in EDs. By employing forecasts, ED resource planners can make informed decisions between balancing demand for ED services and available resources, crucially during outbreaks where ED attendances are high.

The forecast combination incorporated forecast inputs from each forecasting model and provided improvements and consistent performance across different forecast horizons. In addition to the forecast combination, AR and AREV (LASSO) also performed well in short–term forecasting. LASSO allowed for fast selection of predictors and tuning using cross-validation on the training data, which helped to optimize in-sample forecast performance, achieving similar performance to the baseline AR model. RF can achieve similar forecasting performance as the forecast combination in the medium and long–term forecast horizon. Although the GBM model did not perform well in the 1–3 weeks horizon, it performed well in the 6–8 weeks horizon and could outperform forecast combinations. GBM employs an ensemble approach to iteratively decrease the prediction variance resulting from individual weak predictors and as a result, can provide robust forecasts. LSTM-CNN had forecast performance which was worse compared to other sub-models in all forecast horizons (Fig. 3), which may be attributed to insufficient data available to calibrate the LSTM-CNN model. While in certain instances, deep learning models can be trained on a few hundred samples to generate satisfactory results, that is because the trend of the data is not very complex. However, visual inspection reveals that the IPDs ED time series exhibits non-stationarity and drift (Fig. 2A), which are characteristics that are possibly difficult to capture with a few hundred training samples. For deep learning models, a large amount of training data is usually required to get satisfactory performance. This is because these models have many parameters requiring sufficient data to enable effective parameter calibration. In subsequent work, LSTM-CNN can be explored as a methodology to forecast with higher sample size and temporal resolution.

Epidemiological analysis was also performed using LASSO, which could account for the high-dimensional nature of our considered covariates. In this analysis, we found that our conclusions generally aligned well with the related literature in this field. Past studies have shown that higher precipitation is associated with a higher burden of IPDs, such as diarrhea. The incubation period for enteroviruses and bacillary dysentery is 2 to 10 days (about 1 week), which has corroborated with our finding [28]. One possible explanation is that high precipitation can heighten the spread of diarrheal pathogens to humans through contact with polluted water sources [29]. Furthermore, higher temperatures and higher wind speed may reduce the transmission of other infectious diseases [30] with one explanation being that some pathogens may not survive well at warmer temperatures, and higher wind speeds have the potential to dilute the concentration of infectious particles, thereby reducing infection risk [31]. Past studies have also shown that an increase in $PM_{2.5}$ was associated with increased IPD ED attendance [32]. One possible explanation is that infections resulting from $PM_{2.5}$ primarily affect the respiratory system, particularly the lower respiratory tract. This corroborates with other work, which found that lower respiratory infections can be attributed to household PM2.5 pollution, with the highest burdens of PM2.5-attributable lower respiratory infections observed among children under 5 years old and adults over 70 years old [33].

There are several limitations in our study. (**a**) We used data from 2009–2018, a period where the study setting was not subject to periods where structural breaks in transmission occur (i.e. the COVID-19 pandemic period), so model robustness to these phenomena cannot be tested. [34]. Therefore, future work should incorporate data beyond 2020 to understand forecast performance in the presence of structural breaks. (**b**) Additionally, hospital emergency department admissions are rapidly changing and unpredictable during outbreaks, and models which forecast under different temporal resolutions (i.e. daily) should be explored, but outside the scope of this study. (**c**) While the suggested models demonstrated the potential of utilizing information from additional exogenous variables to understand determinants which lead to increased IPDs ED attendances, the interpretation of coefficients was challenging due to the inclusion of numerous predictors and lags selected to train respective model in each forecast window. Ensemble machine learning methods such as RF and GBM do not offer easy means to interpret the impact of exogenous variables on IPDs ED attendances either, as they aggregate predictions from many sub-models. While variable importance plots for ensemble methods can help triangulate influential predictors, it does not provide insights into how each influential predictor affects the disease of interest. The post-selection inference approach was thus used to examine associations between exogenous variables and the outcome of interest. Here, we observed that some estimated coefficients had large absolute values but were not significant after performing post-selection inference. In a multiple regression setting, this situation is likely to be a sign of multicollinearity, i.e., a high correlation between predictor variables. When the predictor variables are highly correlated, the model has difficulty isolating the effect of each predictor, which can inflate the variance of the coefficient estimates, rendering estimates unstable and difficult to interpret. A limitation of LASSO is that when faced with multicollinear predictors, it may select one of them and shrink the other predictors to zero. This means that while LASSO may help provide a more interpretable model in the presence of multicollinearity, it can't alleviate the estimation issue between highly correlated variables. Therefore, it is essential to consider methods that can be effective in avoiding inaccurate estimates due to multicollinearity. (**d**) Only point forecasts were considered for our study. Performing and evaluating density has many advantages over point forecasts [18], as point forecasts cannot describe the associated uncertainty surrounding the forecast. Uncertainty can provide valuable information to policy makers on how confident we are about our forecasts. Ignoring this uncertainty may potentially lead to misinterpretation of forecasts. Therefore, frameworks based on point and density forecasts and forecast evaluation should be established to alleviate these drawbacks. (**e**) Lastly, we only considered the direct forecast strategy and did not apply recursive forecast approach, as the study included many environmental variables that would have imposed a much higher computational burden if a recursive forecast was applied. However, as recursive forecasts were not employed, potentially important dependencies between forecasts could not be modelled. Future work can consider combining recursive and direct forecast strategies, to overcome the inherent limitations of each approach [35].

V. Conclusions

Our proposed forecast combination scheme has the potential to provide forward guidance on short– to long–term healthcare resource allocation and minimize patient wait-time. Its forecast performance remained consistently at the top of all models across a 2-month forecast horizon, with acceptable deterioration rates with increasing forecast horizons.

REFERENCES

[1] K. Sweiss, A. Y. Naser, M. Samannodi, and H. Alwafi, "Hospital admissions due to infectious and parasitic diseases in England and Wales between 1999 and 2019: an ecological study," BMC Infectious Diseases, vol. 22, no. 1, p. 398, Apr. 2022, doi: 10.1186/s12879-022-07388-1.

[2] S. Alonso et al., "The economic burden of malaria on households and the health system in a high transmission district of Mozambique," Malaria Journal, vol. 18, no. 1, p. 360, Nov. 2019, doi: 10.1186/s12936-019-2995-4.

[3] B. Chala and F. Hamde, "Emerging and Re-emerging Vector-Borne Infectious Diseases and the Challenges for Control: A Review," Frontiers in Public Health, vol. 9, 2021, Accessed: Oct. 26, 2023 [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpubh.2021.715759

[4] G. D. Kelen et al., "Emergency Department Crowding: The Canary in the Health Care System," Catalyst non-issue content, vol. 2, no. 5, Sep 2021, doi: 10.1056/CAT.21.0217.

[5] H. C. Oh, W. L. Chow, Y. Gao, L. Tiah, S. H. Goh, and T. Mohan, "Factors associated with inappropriate attendances at the emergency department of a tertiary hospital in Singapore," Singapore Med J, vol 61, no. 2, pp. 75–80, Feb. 2020, doi: 10.11622/smedj.2019041.

[6] F. Harrou, F. Kadri, S. Chaabane, C. Tahon, and Y. Sun, "Improved principal component analysis for anomaly detection: Application to an emergency department," Computers & Industrial Engineering, vol. 88, pp. 63–77, Oct. 2015, doi: 10.1016/j.cie.2015.06.020.

[7] J. E. Hurwitz, J. A. Lee, K. K. Lopiano, S. A. McKinley, J. Keesling, and J. A. Tyndall, "A flexible simulation platform to quantify and manage emergency department crowding," BMC Med Inform Decis Mak, vol. 14, no. 1, p. 50, Jun. 2014, doi: 10.1186/1472-6947-14-50.

[8] Y. Chen, K. Bi, C.-H. (John) Wu, and D. Ben-Arieh, "A new evidence-based optimal control in healthcare delivery: A better clinical treatment management for septic patients," Computers & Industrial Engineering, vol. 137, p. 106010, Nov. 2019, doi: 10.1016/j.cie.2019.106010.

[9] J. Bergs, P. Heerinckx, and S. Verelst, "Knowing what to expect, forecasting monthly emergency department visits: A time-series analysis," International Emergency Nursing, vol. 22, no. 2, pp. 112–115, Apr. 2014, doi: 10.1016/j.ienj.2013.08.001.

[10] F. Kadri, F. Harrou, and Y. Sun, "A multivariate time series approach to forecasting daily attendances at hospital emergency department," in 2017 IEEE Symposium Series on Computational Intelligence (SSCI), Nov. 2017, pp. 1–6. doi: 10.1109/SSCI.2017.8280850.

[11] S. Masum, Y. Liu, and J. Chiverton, "Multi-step Time Series Forecasting of Electric Load Using Machine Learning Models," in Artificial Intelligence and Soft Computing, L. Rutkowski, R. Scherer, M. Korytkowski, W. Pedrycz, R. Tadeusiewicz, and J. M. Zurada, Eds. in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018, pp. 148–159. doi: 10.1007/978-3-319-91253-0_15.

[12] A. R. Khan, K. T. Hasan, T. Islam, and S. Khan, "Forecasting respiratory tract infection episodes from prescription data for healthcare service planning," Int J Data Sci Anal, vol. 11, no. 2, pp 169–180, Mar. 2021, doi: 10.1007/s41060-020-00235-z.

[13] P. Liu, L. Lei, J. Yin, W. Zhang, W. Naijun, and E. El-Darzi, "Healthcare Data Mining: Prediction Inpatient Length of Stay," 2006 3rd International IEEE Conference Intelligent Systems, pp. 832–837, Sep. 2006, doi: 10.1109/IS.2006.348528.

[14] V. K. Sudarshan, M. Brabrand, T. M. Range, and U. K. Wiil, "Performance evaluation of Emergency Department patient arrivals forecasting models by including meteorological and calendar information: A comparative study," Computers in Biology and Medicine, vol. 135, p. 104541, Aug. 2021, doi: 10.1016/j.compbiomed.2021.104541.

[15] F. Harrou, A. Dairi, F. Kadri, and Y. Sun, "Forecasting emergency department overcrowding: A deep learning framework," Chaos, Solitons & Fractals, vol. 139, p. 110247, Oct. 2020, doi: 10.1016/j.chaos.2020.110247.

[16] F. Harrou, A. Dairi, F. Kadri, and Y. Sun, "Effective forecasting of key features in hospital emergency department: Hybrid deep learning-driven methods," Machine Learning with Applications, vol. 7, p 100200, Mar. 2022, doi: 10.1016/j.mlwa.2021.100200.

[17] X. Zhao et al., "A deep learning architecture for forecasting daily emergency department visits with acuity levels," Chaos, Solitons & Fractals, vol. 165, p. 112777, Dec. 2022, doi: 10.1016/j.chaos.2022.112777.

[18] J. T. Lim, E. L. W. Choo, A. Janhavi, K. B. Tan, J. Abisheganaden, and B. Dickens, "Density forecasting of conjunctivitis burden using high-dimensional environmental time series data," Epidemics, vol. 44, p. 100694, Sep. 2023, doi: 10.1016/j.epidem.2023.100694.

[19] J. T. Lim, K. B. Tan, J. Abisheganaden, and B. L. Dickens, "Forecasting upper respiratory tract infection burden using high-dimensional time series data and forecast combinations," PLOS Computational Biology, vol. 19, no. 2, p. e1010892, Feb. 2023, doi: 10.1371/journal.pcbi.1010892.

[20] U. Silachamroon and Y. Wattanagoon, "129 - Paragonimiasis," in Hunter's Tropical Medicine and Emerging Infectious Diseases (Tenth Edition), E. T. Ryan, D. R. Hill, T. Solomon, N. E. Aronson, and T. P. Endy, Eds., London: Elsevier, 2020, pp. 928–931. doi: 10.1016/B978-0-323-55512-8.00129-0.

[21] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[22] E. Hoseinzade and S. Haratizadeh, "CNNpred: CNN-based stock market prediction using a diverse set of variables," Expert Systems with Applications, vol. 129, pp. 273–285, Sep. 2019, doi: 10.1016/j.eswa.2019.03.029.

[23] I. E. Livieris, E. Pintelas, and P. Pintelas, "A CNN–LSTM model for gold price time-series forecasting," Neural Comput & Applic, vol. 32, no. 23, pp. 17351–17360, Dec. 2020, doi: 10.1007/s00521-020-04867-x.

[24] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," Mechanical Systems and Signal Processing, vol. 151, p. 107398, Apr. 2021, doi: 10.1016/j.ymssp.2020.107398.

[25] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization." arXiv, Jan. 29, 2017. doi: 10.48550/arXiv.1412.6980.

[26] G. Marcjasz, B. Uniejewski, and R. Weron, "Beating the Naïve—Combining LASSO with Naïve Intraday Electricity Price Forecasts," Energies, vol. 13, no. 7, Art. no. 7, Jan. 2020, doi: 10.3390/en13071667.

[27] F. X. Diebold and R. S. Mariano, "Comparing Predictive Accuracy," Journal of Business & Economic Statistics, vol. 20, no. 1, pp. 134–144, Jan. 2002, doi: 10.1198/073500102753410444.

[28] M.-J. Chen, C.-Y. Lin, Y.-T. Wu, P.-C. Wu, S.-C. Lung, and H.-J. Su, "Effects of Extreme Precipitation to the Distribution of Infectious Diseases in Taiwan, 1994–2008," PLOS ONE, vol. 7, no. 6, p. e34651, Jun. 2012, doi: 10.1371/journal.pone.0034651.

[29] G. Andhikaputra et al., "The impact of temperature and precipitation on all-infectious-, bacterial-, and viral-diarrheal disease in Taiwan," Sci Total Environ, vol. 862, p. 160850, Mar. 2023, doi: 10.1016/j.scitotenv.2022.160850.

[30] S. A. P. Clouston, O. Morozova, and J. R. Meliker, "A wind speed threshold for increased outdoor transmission of coronavirus: an ecological study," BMC Infect Dis, vol. 21, p. 1194, Nov. 2021, doi: 10.1186/s12879-021-06796-z.

[31] N. Islam, S. Shabnam, and A. M. Erzurumluoglu, "Temperature, humidity, and wind speed are associated with lower Covid-19 incidence." medRxiv, p. 2020.03.27.20045658, Mar. 31, 2020. doi: 10.1101/2020.03.27.20045658.

[32] M. Szyszkowicz and N. de Angelis, "Air Pollution and Emergency Department Visits for Infectious Diseases in Toronto, Canada." Preprints, Jan. 25, 2022. doi: 10.20944/preprints202201.0388.v1.

[33] X. Bu et al., "Global PM2.5-attributable health burden from 1990 to 2017: Estimates from the Global Burden of disease study 2017," Environmental Research, vol. 197, p. 111123, Jun. 2021, doi: 10.1016/j.envres.2021.111123.

[34] J. Ditzen, Y. Karavias, and J. Westerlund, "Testing and Estimating Structural Breaks in Time Series and Panel Data in Stata." arXiv, Oct. 28, 2021. doi: 10.48550/arXiv.2110.14550.

[35] N. H. An and D. T. Anh, "Comparison of Strategies for Multi-step-Ahead Prediction of Time Series Using Neural Network," 2015 International Conference on Advanced Computing and Applications (ACOMP), pp. 142–149, Nov. 2015, doi: 10.1109/ACOMP.2015.24.