Chaos, Solitons and Fractals 168 (2023) 113170

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

journal homepage: www.elsevier.com/locate/chaos

A deep learning based hybrid architecture for weekly dengue incidences forecasting

Xinxing Zhao [a], Kainan Li [a], Candice Ke En Ang [b], Kang Hao Cheong [a,*]

[a] Science, Mathematics and Technology Cluster, Singapore University of Technology and Design, 8 Somapah Road, S487372, Singapore
[b] MOH Holdings Pte Ltd, 1 Maritime Square, S099253, Singapore

ARTICLE INFO

Keywords:
Epidemiology
Time series forecasting
Dengue incidences forecasting
Deep learning
Hybrid models

ABSTRACT

Dengue is a mosquito-borne viral disease widely spread in tropical and subtropical regions. Its adverse impact on the human health and global economies cannot be overstated. In order to implement more effective vector control measures, mechanisms that can more accurately forecast dengue cases are needed more urgently than before. In this paper, a novel hybrid architecture which has the advantages of both convolutional neural networks and recurrent neural networks is being proposed to forecast weekly dengue incidence. The forecasting performance of this architecture reveals that the deep hybrid architecture outperforms other frequently used deep learning models in dengue forecasting tasks. We have also evaluated the proposed models against state-of-the-art studies in the literature, demonstrating that our proposed hybrid models utilizing recurrent networks with convolutional layers can provide a significant boost in dengue forecasting.

1. Introduction

Dengue fever is an acute mosquito-borne viral disease common in tropical and subtropical regions. This disease affects 100–400 million [1] people yearly, and the clinical manifestations include nausea, vomiting, rash and body aches [2]. In severe cases, dengue may cause plasma leakage, leading to circulatory failure and clinically dengue shock syndrome or even death [3]. Currently, the mainstay of treatment for dengue is supportive, and there is yet to be an effective cure developed [4]. Vaccination for the dengue virus is still a work in progress due to its complexity with four viral serotypes [5], and are currently inadequate in stopping the spread of the disease. Therefore, elimination of the infected Aedes mosquito remains the most effective method in suppressing transmission of the disease. Nevertheless, the tedious task of vector control may be resource and labour intensive without novel and innovative methods, and therefore poses a huge economic burden in the affected regions. To utilize resources more effectively, a possible strategy would be to accurately forecast dengue cases early and subsequently plan and allocate resources accordingly, such as deploying more mosquito traps or using Wolbachia-Aedes suppression technology [6] in the region.

The complex interplay between epidemiological and environmental determinants of dengue causes the understanding of the long-term epidemiological trends of the disease to be difficult to grasp, limiting our ability to predict the disease incidence. There are usually seasonal variations of outbreaks in certain regions, where the epidemic season fluctuates, possibly contributed by the immunological and demographical structures of the population, as well as environmental and weather factors.

There are different models for forecasting dengue incidence in the literature, ranging from statistical models such as autoregressive models [7] and regression models [8], which correlate dengue cases with factors such as environmental data and time lags, to machine learning models such as support vector machine [9], decision tree and k-nearest neighbor [10] models, which find patterns from historical dengue cases and learns the non-linear relationship between historical data and the cases being forecast. Artificial intelligence approaches [11,12] have become more prevalent due to their efficiency, accuracy and flexibility over other models [13,14]. It is also worth noting that hybrid models formed by combining different approaches [15,16] may achieve better accuracies in time series forecasting problems.

Recently, deep learning models have been successfully in achieving excellent results in many challenging real-world forecasting [17,18] and classifying [19,20] problems. RNN, LSTM and GRU models are frequently used, and they are better suited for analyzing temporal and sequential data. They have been used in many domains such as forecasting patient visits at emergency department [21], traffic flow [22] in the intelligent transportation system, building energy consumption [23], tourism demand [24], and the Covid-19 pandemic transmission [25].

In this paper, a hybrid architecture which has recurrent neural network related parts (RNN, GRU, LSTM and BiLSTM) and CNN parts

* Corresponding author.
E-mail address: kanghao_cheong@sutd.edu.sg (K.H. Cheong).

https://doi.org/10.1016/j.chaos.2023.113170
Received 22 November 2022; Received in revised form 13 January 2023; Accepted 18 January 2023
Available online 6 February 2023

as components was proposed and applied to forecast weekly dengue cases in Singapore. Using hybrid architecture is novel in the context of forecasting dengue cases, as we stack two types (hybrid) of deep learning models in dengue incidence forecasting, and this research is inspired by a previous work [26] on emergency department visitors forecasting. The rest of this paper is organized as follows. Section 2 gives the related work and Section 3 presents the methodologies. Section 4 provides results followed by the discussion in Sections 5 and 6 concludes this study.

2. Related work

Applying deep learning techniques in dengue incidence forecasting is a recent development in the research community. Past studies [21, 27] have demonstrated their strong forecasting abilities as compared to that of statistical and traditional machine learning models in time series forecasting problems. Hence, we will be focusing more on deep learning related models, which will be discussed in the following sections.

Xu et al. [11] used a LSTM model with transfer learning to forecast dengue cases in 20 cities in China. They compared their proposed models with BPNN, GAM, SVR and GBM models, and concluded that their models have better performance in terms of RMSE in most of the cases. However, the monthly dengue cases in their studied time period were quantitatively very small. Taking Guangzhou as an example, which was the most seriously impacted city in the region by dengue [28] in that period of time, the number of monthly cases varied largely, and in many months there were few or even no dengue cases at all. This could have greatly affected the forecasting accuracy in their research. Mussumeci et al. [29] used LSTM and Random Forest regression models with temperature, humidity, atmospheric pressure and tweets about the dengue cases as features for dengue weekly forecasting in 790 Brazilian cities. They concluded that LSTM models can be used in large scale epidemiological forecasting problems. Nguyen et al. [30] used LSTM, attention-enhanced LSTM, CNN and traditional machine learning models with meteorological factors to predict monthly dengue cases in 20 provinces of Vietnam. They concluded that deep learning models are better than traditional machine learning models in the forecasting problem, and that attention-enhanced LSTM is best model among deep learning models (in terms of both MAE and RMSE). Sanchez et al. [31] used a LSTM model for dengue time series forecasting with ovitrap data as a predictor, and it outperformed traditional machine learning mechanisms such as Random Forest and Lasso regression. The models achieved a correlation coefficient of 0.92. Saleh et al. [32] developed a LSTM model and compared it to a SVR model with weather and climate data. The results indicate that LSTM performed better than SVR, with R^2 scoring 0.75 and also better captured the rising and falling trends of dengue cases. Chakraborty et al. [15] proposed a hybrid architecture of ARIMA and NNAR (ANN is a form of deep learning and NNAR is one kind of ANN's) models for dengue forecasting. They tested the hybrid architecture in dengue cases forecasting in three dengue endemic regions, and it achieved better forecasting performance in terms of RMSE, MAE and SMAPE (best 0.57) than ARIMA, SVM, ANN, LSTM, NNAR as well as hybrid ARIMA-SVM, ARIMA-ANN and ARIMA-LSTM models in monthly dengue cases forecasting.

Currently, the LSTM model is the most frequently used deep learning technique in dengue cases forecasting problems and the performance of deep learning models are usually better than more traditional machine learning and statistical models such as Lasso, Random Forest, SVR, ARIMA. Performance metrics such as RMSE, MAE and R^2 are often used. However, the forecasting accuracy in these research are not always reported, as dengue incidence are hard to forecast and resulting accuracies are not as good as those derived from other time series related problems. Despite the above, some studies had reported their forecasting accuracies as discussed in the following. Chakraborty et al. [15] used a hybrid model combining an ARIMA and an autoregressive model for monthly dengue cases forecasting, with the model's best performance of 57% in terms of SMAPE for 3 months in advance forecasting. Withanage et al. [8] used regression models for monthly dengue cases forecasting, and their best model achieved a MAPE of 39.37%. Shi et al. [33] used a Lasso model with meteorological data to forecast the weekly incidence of dengue in Singapore, and reported they achieved a MAPE of 17% (therefore forecasting accuracy is 83%) for the immediate upcoming week forecasting. In terms of prior literature, 83% forecasting accuracy appeared to be the best performance for similar problem statements involving weekly dengue cases forecasting. Our study aims to provide a deep learning framework which can improve state-of-the-art forecasting accuracy in the literature.

3. Methods

In this section, we first introduce the data set of weekly dengue cases over 10 years, and provide details of the proposed hybrid architecture. Next, we discuss the configuration details for different deep learning models. Finally, we present the evaluation metrics used for assessing performance for comparison between the different models.

3.1. Data collection

The weekly reported dengue cases were obtained from the Singapore Ministry of Health website (https://www.moh.gov.sg/resources-statistics/infectious-disease-statistics/2022/weekly-infectious-diseases-bulletin), and the data used for our analysis covers a 10-year period, from week 24 (10–16 June) of 2012 to week 21 (15–21 May) of 2022 (for years other than 2022, just replace 2022 with the other year in the MOH hyperlink). Fig. 1 demonstrates the weekly dengue cases of this data set, with added information of average temperature and rainfall included. Notably, there were a few peaks in the past 10 years. The most serious dengue outbreak was in 2020, where there were more than 600 infected cases a week, and this had lasted about 20 weeks consecutively, with the peak weekly cases at about 1800. The x-axes represents the weeks over the ten years, while the y-axes are the dengue cases in the same period. Average weekly temperature in °C and weekly rainfall in mm are also provided in the same figure.

3.2. The proposed architecture

RNN models are frequently used with time series data, similar to feedforward and convolutional neural networks, where they learn from training data. However, they are distinguished by their memory capabilities (calculated from previous inputs and internal states) which can influence the current input and output. As past states of a network are helpful in determining the output of a future event, RNN models generally have a better prediction performance than the unidirectional recurrent neural networks which assume inputs and outputs are independent. RNN has variants such as LSTM [34] and GRU [25]; The Bidirectional LSTM networks (BiLSTMs) [35] are developments of LSTMs in which additional training will be enabled by traversing the input data via two directions.

CNN models are frequently used with image processing related problems, such as in classification [36], segmentation [37] and recognition [38]. Recently, they have also been adapted successfully in many time series related problems such as forecasting of COVID-19 cases [39], financial time series [40], and the number of patient visits to the Emergency Department [26].

One of the key abilities of CNN models is that they can automatically find key features from a data set, while RNN and its variants are good at dealing with tasks that involve sequential inputs. Henceforth, a deep hybrid architecture combining these two types of deep learning models is being proposed, as shown in Fig. 2. It has mainly two parts, the convolutional and recurrent modules. The weekly dengue cases and climate features are the input for the architecture, and after passing through the CNN and Recurrent modules, the processed signals will be

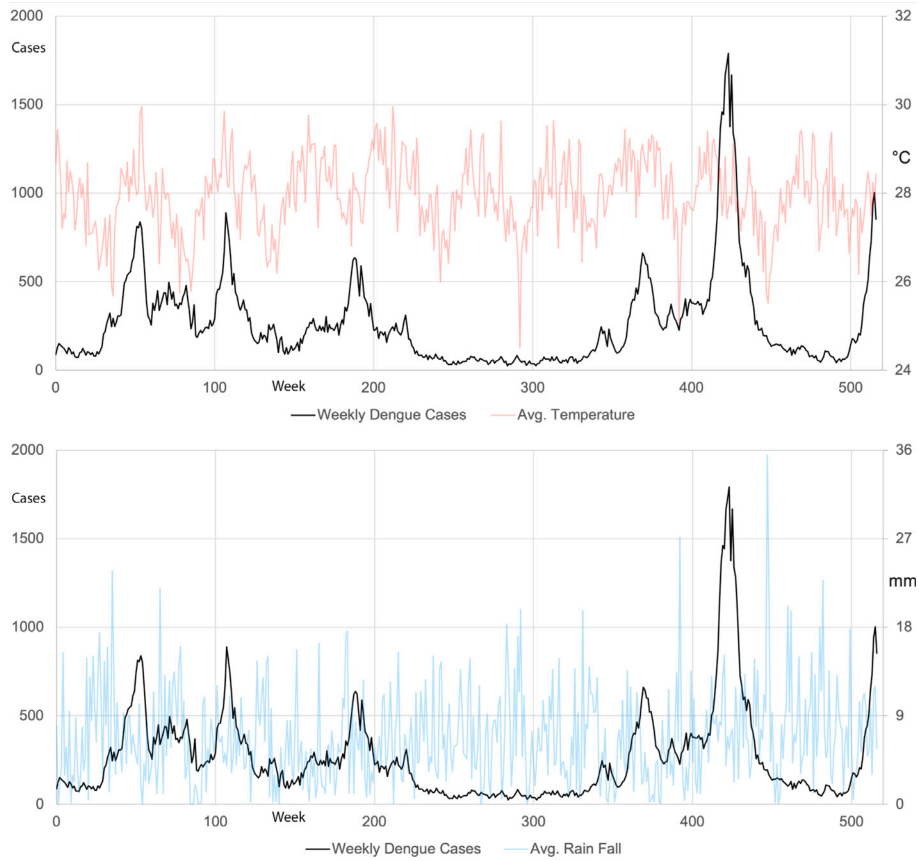

Fig. 1. The weekly dengue cases, average temperature and rainfall in Singapore (2012–2022).

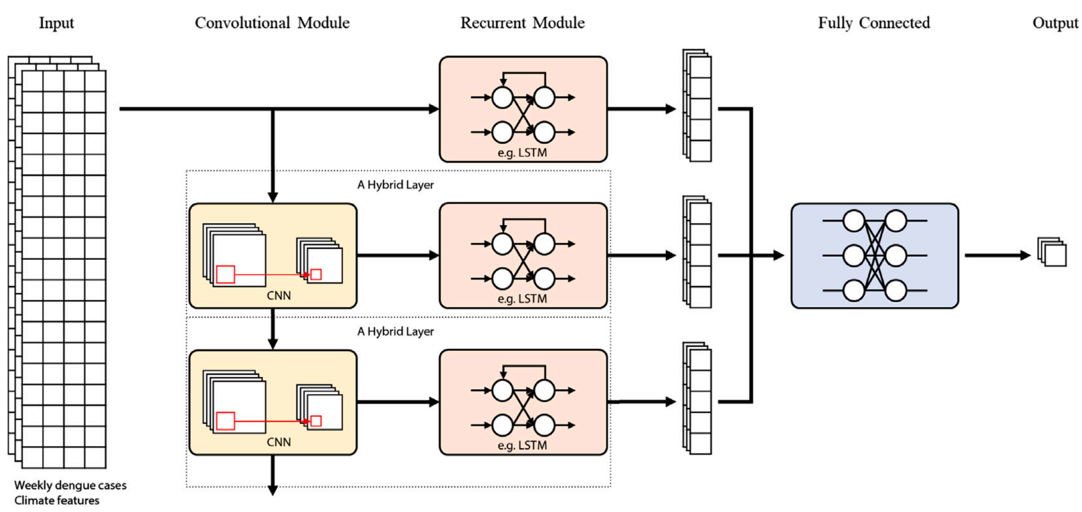

Fig. 2. The proposed architecture for dengue cases forecasting.

concatenated, subsequently passing through two fully connected layers before giving the forecasting output. More CNN and Recurrent modules combined hybrid layers (as highlighted in the box of Fig. 2) can be added for the structure to go deeper. We call a convolutional module and a recurrent module in the dotted box a stack (also known as a hybrid layer). In the implementation stage, we tested the model with different numbers of stacks and used RNN, GRU, LSTM and BiLSTM separately as the recurrent modules. For example, when LSTM is used as the recurrent modules, only LSTM was used for all the Recurrent modules in the architecture.

Different forecasting horizons in the dengue incidence forecasting problems are available in the literature, such as seasonally [9], monthly [15] and weekly [33]. In this study, we forecast weekly and in steps of 1, 2, 3 and 4 weeks into the future, as this will likely allow for adequate assessment and planning of required resources by healthcare agencies to deal with situations such as outbreaks.

3.3. Configurations, settings and features

In the proposed architecture, each CNN module has a convolutional 2D part, with kernel size 3 and padding 1. For each recurrent module in

Table 1
Configurations for the modules in the architecture.

Module	Parameters & settings
Recurrent Module	3 layers, each with 50 hidden units, dropout rate 0.25, activation function ReLu.
CNN Module	Convolutional 2D, kernel size (3, 1) padding (1, 0).

Table 2
Configurations for the benchmark models.

Model	Parameters & settings
RNN and variants	3 layers, each with 50 hidden units, dropout rate 0.25, activation function ReLu.
CNN model	3 convolutional layers, kernel size 2 and stride 2, dropout rate of 0.25, activation function ReLu.
ConvLSTM model	3 convolutional layers, kernel size 3, padding 1, 2 max-pooling layers, LSTM part has 3 layers, 50 hidden units each, dropout rate 0.25, activation function ReLu.

Table 3
System settings.

Experimental platform	Information
Software environment	Python: 3.9; PyTorch: 1.10.1.
Hardware environment	Operating System: Window 10.0 64-bit; CPU: Intel(R) Core(TM) i7-8750H @ 2.21 GHz GPU: NVIDIA GeForce GTX 1060 Hard Disk Capacity: 1 TB Memory: 16 GB

the architecture, RNN and its variants (GRU, LSTM and BiLSTM models) are used with only one model type each time. For example, as shown in Fig. 2, if the LSTM model is used, all the recurrent modules are with LSTM models. For each recurrent module in the proposed architecture and for LSTM, BiLSTM, RNN, GRU models (benchmark models), they share the same configurations, 3 layers, 50 hidden units in each layer, dropout rate is 0.25, and activation function is ReLu.

For the benchmark CNN, it has 3 convolutional layers, each layer with kernel size 2 and stride 2, dropout rate of 0.25, ReLu as the activation function. For the benchmark ConvLSTM, it has 3 convolutional layers with kernel size 3, padding 1, 2 max-pooling layers, and LSTM part has 3 layers, 50 hidden units each, dropout rate is 0.25 in each layer, and ReLu as the activation function. These configurations are listed in Tables 1 and 2 for greater clarity.

For the proposed architecture and benchmark models, the chosen parameters and settings were adapted from previous deep learning research in time series forecasting [17,18,21,26] which have demonstrated effectiveness in their respective problems. In the current study, these settings were tested and have achieved very good forecasting performances.

The 10-year data set was divided into training and testing sets, with 70% used for training, and the remaining 30% used for testing (to avoid the overfitting problem). This is similar to many other time series related forecasting research [17,27,41,42], with dengue incidence forecasting included.

For the features, we have used weekly dengue cases, mean temperature and rainfall. The details of the system settings are listed in Table 3.

3.4. Evaluation metrics

Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) are used to quantify the forecasting performance of different deep learning models, where n corresponds to the number of weeks in the study, y_t is the reported cases of dengue at week t, and $\hat{y}_t$ corresponds to the forecast cases at week t.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2} \tag{1}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |\ (y_t - \hat{y}_t)\ | \tag{2}$$

$$MAPE = \frac{100\%}{n}\sum_{t=1}^{n}\frac{|\ (y_t - \hat{y}_t)\ |}{y_t} \tag{3}$$

4. Results

Figs. 3 and 4 provide the performance of hybrid models and non-hybrid deep learning models in terms of MAPE, MAE and RMSE with different time steps, as well as with and without weather (climate) features. Future steps are labeled 1, 2, 3 and 4, corresponding to 1 to 4 weeks of forecast into the future.

For 1 week in advance forecasting, for MAPE performance: Benchmark models are in the range of 12.70–13.83% (therefore forecasting accuracy in the range of 86.17–87.30%) without weather features and 12.73–14.77% (accuracy 85.23–87.27%) with weather features. Hybrid deep learning models are in the range of 12.28–14.14% (accuracy 85.86–87.72%) without weather features, and 12.48–13.95% (accuracy 86.05–87.52%) with weather features. The best forecasting accuracy is 87.72%, achieved by 1 stack of CNN-BiLSTM model without weather features. For MAE performance: Benchmark models are in the range of 41.21–48.12 and 42.81–52.68 without and with weather features respectively; Hybrid deep learning models are in the range of 41.40–49.97 and 42.35–47.61 without and with weather features respectively. The best MAE is 41.21, achieved by RNN model, and 1 stack CNN-BiLSTM model achieved close performance of 41.40, both without weather features. For RMSE performance: Benchmark models are in the range of 69.25–88.24 and 75.08–88.98 without and with weather features respectively. Hybrid deep learning models are in the range of 73.30–85.00 and 70.59–85.90 without and with weather features respectively. The best RMSE is 69.25, achieved by RNN model without weather features.

For 2 weeks in advance forecasting, for MAPE performance: Benchmark models are in the range of 16.35–17.78% without weather features and 16.74–18.13% with weather features. Hybrid deep learning models are in the range of 15.39–17.89% without weather features, and 16.23–18.10% with weather features. The best forecasting accuracy is 84.61% (1-MAPE), achieved by 1 stack of CNN-BiLSTM model without weather features. For MAE performance: Benchmark models are in the range of 56.52–65.15 and 58.15–66.73 without and with weather features respectively. Hybrid deep learning models are in the range of 53.01–68.15 and 56.48–67.10 without and with weather features respectively. The best MAE is 53.01, achieved by 1 stack of CNN-BiLSTM model without weather features. For RMSE performance: Benchmark models are in the range of 94.64–113.79 and 96.29–111.60 without and with weather features respectively. Hybrid deep learning models are in the range of 90.33–121.48 and 94.46–115.09 without and with weather features respectively. The best RMSE is 90.33, achieved by 1 stack of CNN-BiLSTM model without weather features.

For 3 weeks in advance forecasting, for MAPE performance: Benchmark models are in the range of 19.82–22.40% without weather features and 19.73–22.35% with weather features. Hybrid deep learning models are in the range of 18.51–21.65% without weather features and 19.54–22.46% with weather features. The best forecasting accuracy is 81.49%, achieves by 1 stack of CNN-BiLSTM model without weather features. For MAE performance: Benchmark models are in the range of 73.39–82.48 and 72.00–85.97 without and with weather features respectively. Hybrid deep learning models are in the range of 65.99–85.55 and 70.19–85.11 without and with weather features respectively. The best MAE is 65.99, achieved by 1 stack of CNN-BiLSTM model

Future Steps	Model	MAPE		MAE		RMSE	
		w/o weather	w/ weather	w/o weather	w/ weather	w/o weather	w/ weather
1	Benchmark: BiLSTM	0.13368	0.13019	45.013	45.577	75.325	79.881
	Benchmark: CNN	0.13505	0.13279	44.883	45.758	73.006	76.797
	Benchmark: ConvLSTM	0.13825	0.14773	47.126	52.683	78.277	88.642
	Benchmark: GRU	0.13728	0.14743	48.182	50.398	88.237	88.977
	Benchmark: LSTM	0.13683	0.13851	47.433	47.499	80.961	79.915
	Benchmark: RNN	0.12701	0.12730	41.213	42.812	69.250	75.079
	CNN-BiLSTM (1 stack)	0.12275	0.12938	41.397	43.646	75.571	75.941
	CNN-BiLSTM (2 stacks)	0.13070	0.13053	44.428	46.062	82.549	77.349
	CNN-BiLSTM (3 stacks)	0.12556	0.13146	42.371	45.796	77.481	79.732
	CNN-GRU (1 stack)	0.13595	0.13951	48.984	46.317	83.072	85.894
	CNN-GRU (2 stacks)	0.13164	0.13471	46.183	44.675	81.406	74.650
	CNN-GRU (3 stacks)	0.14136	0.13627	49.972	46.943	83.276	83.994
	CNN-LSTM (1 stack)	0.13646	0.13505	48.634	47.608	85.001	79.197
	CNN-LSTM (2 stacks)	0.13152	0.13288	46.068	46.875	80.045	81.404
	CNN-LSTM (3 stacks)	0.13085	0.12874	45.635	42.735	79.781	78.544
	CNN-RNN (1 stack)	0.13172	0.12702	44.311	42.353	73.301	70.592
	CNN-RNN (2 stacks)	0.12891	0.12483	44.598	42.398	77.183	73.163
	CNN-RNN (3 stacks)	0.12627	0.12923	43.518	44.914	73.911	77.736
2	Benchmark: BiLSTM	0.16759	0.16779	60.016	58.489	99.462	98.900
	Benchmark: CNN	0.17442	0.17834	61.436	61.256	99.962	102.264
	Benchmark: ConvLSTM	0.17716	0.18075	63.148	66.727	103.998	110.960
	Benchmark: GRU	0.17206	0.18127	63.562	65.110	113.789	111.597
	Benchmark: LSTM	0.17783	0.18029	65.151	63.574	110.851	110.112
	Benchmark: RNN	0.16349	0.16735	56.521	58.151	94.639	96.290
	CNN-BiLSTM (1 stack)	0.15387	0.16231	53.007	56.478	90.326	94.458
	CNN-BiLSTM (2 stacks)	0.16743	0.16730	60.479	62.270	103.406	106.976
	CNN-BiLSTM (3 stacks)	0.16270	0.16581	59.516	58.905	107.166	95.655
	CNN-GRU (1 stack)	0.17431	0.18098	63.726	62.451	112.218	106.394
	CNN-GRU (2 stacks)	0.16992	0.17298	64.151	62.477	116.275	106.099
	CNN-GRU (3 stacks)	0.17892	0.17544	68.145	63.161	121.479	115.087
	CNN-LSTM (1 stack)	0.17815	0.17988	65.601	67.102	116.006	113.605
	CNN-LSTM (2 stacks)	0.17419	0.17545	64.315	64.486	111.573	111.134
	CNN-LSTM (3 stacks)	0.17196	0.17109	62.480	61.226	106.861	103.531
	CNN-RNN (1 stack)	0.17244	0.17381	60.969	61.535	101.617	104.681
	CNN-RNN (2 stacks)	0.16742	0.16336	60.412	56.876	104.114	95.847
	CNN-RNN (3 stacks)	0.16721	0.17440	61.394	61.534	104.216	105.074

Fig. 3. The dengue forecasting performance of deep learning models with steps of 1 and 2 weeks.

Future Steps	Model	MAPE		MAE		RMSE	
		w/o weather	w/ weather	w/o weather	w/ weather	w/o weather	w/ weather
3	Benchmark: BiLSTM	0.19816	0.19733	73.386	72.004	124.934	122.563
	Benchmark: CNN	0.21302	0.21589	78.336	77.082	131.476	133.737
	Benchmark: ConvLSTM	0.21156	0.21877	78.226	81.630	130.816	136.715
	Benchmark: GRU	0.21387	0.22043	81.227	85.968	148.345	151.941
	Benchmark: LSTM	0.22401	0.22346	82.483	84.251	141.648	149.732
	Benchmark: RNN	0.20259	0.20546	73.727	73.737	126.594	126.384
	CNN-BiLSTM (1 stack)	0.18507	0.19544	65.987	70.186	112.648	118.506
	CNN-BiLSTM (2 stacks)	0.20520	0.20664	75.877	79.844	130.284	139.503
	CNN-BiLSTM (3 stacks)	0.19782	0.20029	76.488	73.150	139.277	120.439
	CNN-GRU (1 stack)	0.21075	0.22460	79.979	79.757	145.807	139.287
	CNN-GRU (2 stacks)	0.20741	0.20601	82.764	78.272	154.936	139.882
	CNN-GRU (3 stacks)	0.21507	0.22348	85.547	81.193	156.044	141.416
	CNN-LSTM (1 stack)	0.21520	0.22351	82.194	85.111	153.941	145.885
	CNN-LSTM (2 stacks)	0.20891	0.21725	80.301	79.000	141.519	136.393
	CNN-LSTM (3 stacks)	0.21659	0.21200	79.201	74.842	136.277	128.460
	CNN-RNN (1 stack)	0.21161	0.22044	77.972	79.856	133.401	137.276
	CNN-RNN (2 stacks)	0.21307	0.20157	78.100	71.450	137.520	121.378
	CNN-RNN (3 stacks)	0.20840	0.21732	79.169	78.703	136.046	133.081
4	Benchmark: BiLSTM	0.22955	0.22561	85.947	85.115	148.364	145.669
	Benchmark: CNN	0.25022	0.25484	94.108	92.204	158.588	162.501
	Benchmark: ConvLSTM	0.24520	0.25695	92.687	96.017	156.540	161.382
	Benchmark: GRU	0.25596	0.25596	97.385	105.417	178.947	185.934
	Benchmark: LSTM	0.26274	0.26224	99.468	100.887	171.071	184.868
	Benchmark: RNN	0.23939	0.24690	87.898	89.840	150.203	155.360
	CNN-BiLSTM (1 stack)	0.21723	0.22899	79.439	84.054	136.355	142.979
	CNN-BiLSTM (2 stacks)	0.24033	0.24330	90.586	93.902	156.036	165.799
	CNN-BiLSTM (3 stacks)	0.22615	0.23356	86.221	87.049	152.394	143.133
	CNN-GRU (1 stack)	0.24707	0.26772	95.820	95.888	171.511	168.234
	CNN-GRU (2 stacks)	0.24690	0.23493	101.694	90.710	191.119	163.956
	CNN-GRU (3 stacks)	0.24961	0.25458	97.874	94.430	172.219	165.834
	CNN-LSTM (1 stack)	0.25038	0.27028	96.658	102.988	177.827	177.216
	CNN-LSTM (2 stacks)	0.24248	0.25764	95.862	94.647	172.100	165.979
	CNN-LSTM (3 stacks)	0.25437	0.25154	94.644	88.422	162.601	155.364
	CNN-RNN (1 stack)	0.25493	0.26033	93.130	95.104	159.994	163.760
	CNN-RNN (2 stacks)	0.26053	0.23744	95.754	85.588	162.994	145.995
	CNN-RNN (3 stacks)	0.24932	0.25436	95.918	94.012	166.501	158.183

Fig. 4. The dengue forecasting performance of deep learning models with steps of 3 and 4 weeks.

without weather features. For RMSE performance: Benchmark models are in the range of 124.93–148.35 and 122.56–151.94 without and with weather features respectively. Hybrid deep learning models are in the range of 112.65–156.04 and 118.51–145.89 without and with weather features respectively. The best RMSE is 112.65, achieved by 1 stack of CNN-BiLSTM model without weather features.

For 4 weeks in advance forecasting, for MAPE performance: Benchmark models are in the range of 22.96–26.27% without weather features and 22.56–26.22% with weather features. Hybrid deep learning models are in the range of 21.72–26.05% without weather features and 22.90–27.03% with weather features. The best forecasting accuracy is 78.28%, achieved by 1 stack of CNN-BiLSTM model without weather features. For MAE performance: Benchmark models are in the range of 85.95–99.47 and 85.12–105.42 without and with weather features respectively. Hybrid deep learning models are in the range of 79.44–101.69 and 84.05–102.99 without and with weather features respectively. The best MAE is 79.44, achieved by 1 stack of CNN-BiLSTM model without weather features. For RMSE performance: Benchmark

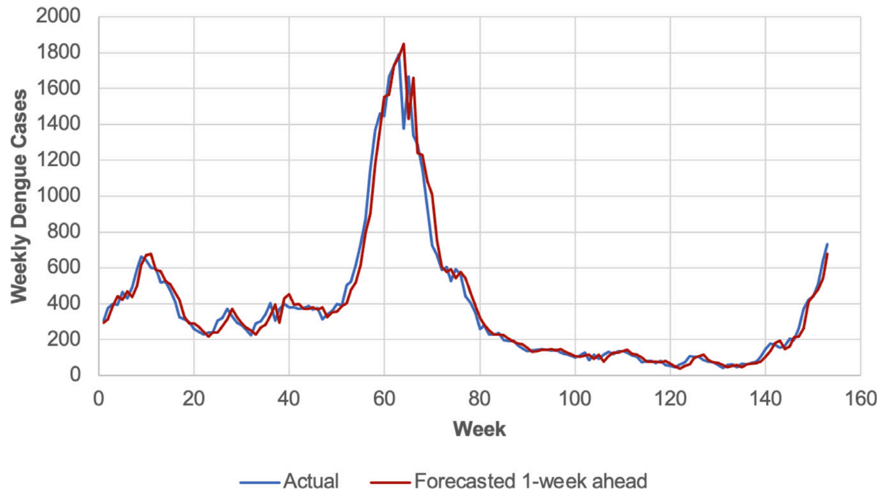

Fig. 5. The dengue forecasting performance with CNN-BiLSTM model for 1 week in advance.

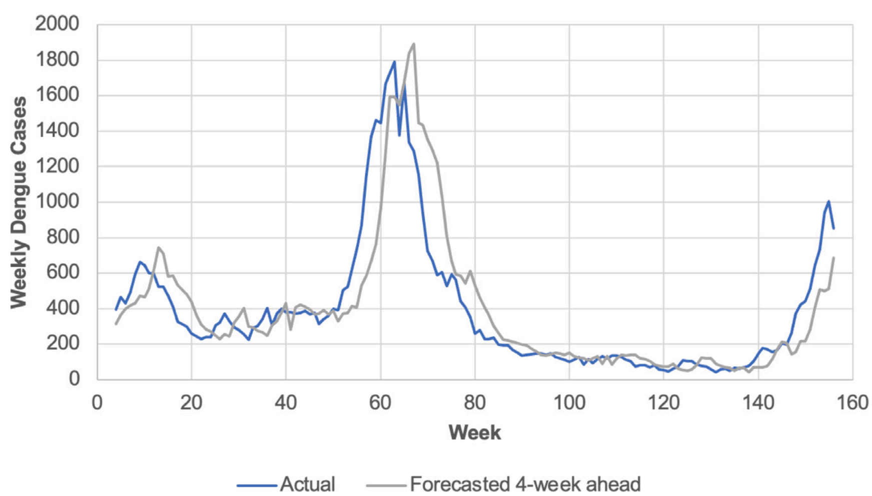

Fig. 6. The dengue forecasting performance with CNN-BiLSTM model for 4 weeks in advance.

models are in the range of 148.36–178.95 and 145.67–185.93 without and with weather features respectively. Hybrid deep learning models are in the range of 136.36–191.12 and 142.98–177.22 without and with weather features respectively. The best RMSE is 136.36, achieved by 1 stack of CNN-BiLSTM model without weather features.

Figs. 5 and 7 provide visual performance for the best hybrid deep learning model (1 stack CNN-BiLSTM) for 1 week in advance forecasting. Notably, the hybrid deep learning model achieved very good performance and can predict the trend of weekly cases well. Figs. 6 and 8 provide visual performance for the best hybrid deep learning model (still the 1 stack CNN-BiLSTM) for 4 weeks in advance forecasting. We observe that the deep learning model still achieved reasonable forecasting performance. Performance figures for 2 and 3 weeks in advance forecasting will not be provided, as it appears that 2 weeks in advance forecasting has close performance to 1 week forecasting, while 3 weeks in advance forecasting are close to both 2 weeks and 4 weeks in advance forecasting weeks, as can be seen from Figs. 3 and 4. Fig. 9 gives the runtime vs. MAPE for different deep learning models for 1 week in advance forecasting without weather features, while other scenarios (2, 3 and 4 weeks) in advance forecasting with or without weather features have similar runtime for different deep learning models, therefore we do not provide them here. As we can see from the figure, the benchmark models have lesser runtime than these stacked hybrid models, and the hybrid models have more runtime when the architecture is deeper with more hybrid stacks added. The runtime for the different models are in $[1000\ \mathrm{ms}, 9000\ \mathrm{ms}]$ range.

5. Discussion

Using deep learning techniques for forecasting dengue cases is a relatively new research area. As mentioned in Section 2, LSTM is one of the most frequently used deep learning models, while other deep learning models such as CNN and ANN are also adapted and used in this area. In this study, we use a hybrid structure which stacks deep recurrent layers (RNN, LSTM, GRU, BiLSTM) and CNN layers for forecasting weekly dengue incidence in Singapore.

As demonstrated in Figs. 3 and 4, the developed deep learning hybrid models achieve good performance. MAPE is one of the most intuitive metrics for the performance of the deep learning models, as it can give the exact level of accuracy (1-MAPE is the accuracy) and we pay more attention to it. The best forecasting accuracy achieved for forecasting 1 week in advance in this study is 87.72%. In current literature, the best weekly forecasting accuracy was obtained by Shi et al. [33], who used a Lasso model with weather factors for dengue incidence forecasting and achieved a MAPE of 17% for 1 week forecasting. We have achieved better weekly dengue forecasting with our model, outperforming their model by 4.72% in terms of accuracy. It is also worth noting that in Shi et al. amongst the 12-year weekly

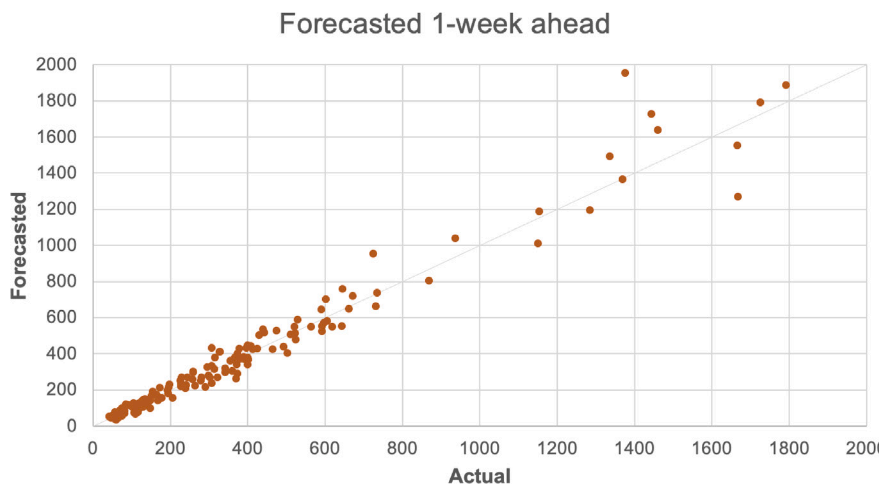

Fig. 7. The scatter plot for dengue forecasting performance with CNN-BiLSTM model for 1 week in advance.

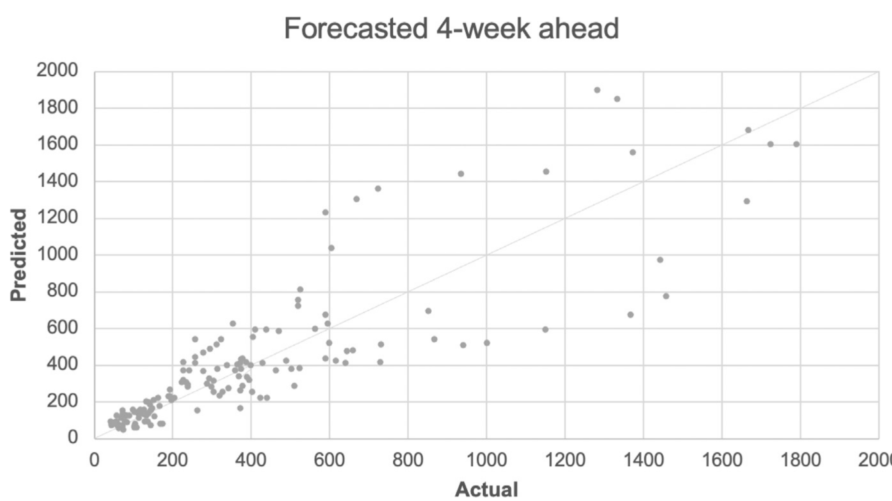

Fig. 8. The scatter plot for dengue forecasting performance with CNN-BiLSTM model for 4 weeks in advance.

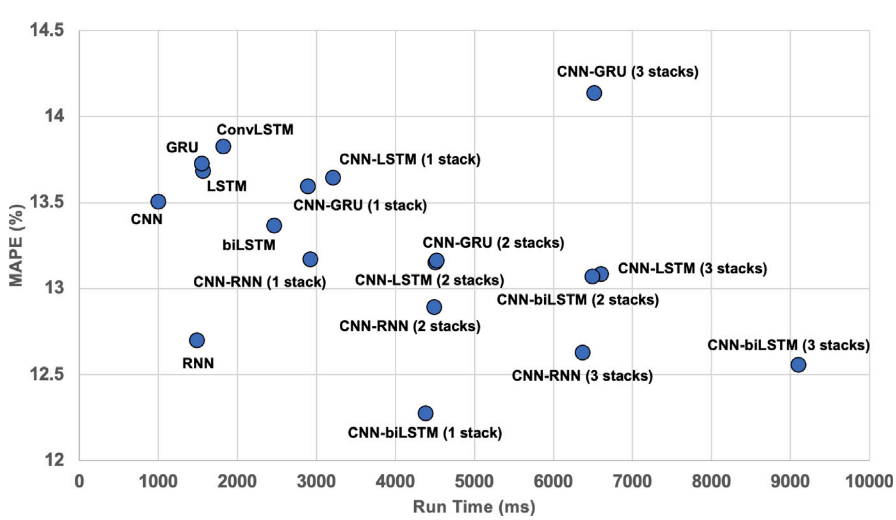

Fig. 9. Runtime vs. MAPE for deep learning models (1 week in advance forecasting).

data, 10-year data was used as the training set and 2-year as the testing set,potentiating the overfitting problem. For 2, 3 and 4 weeks in advance forecasting, the performance of the developed models in this study are also good, with the best accuracy achieved at 84.61%, 81.49%, 78.28% respectively.

The performance of these hybrid models are also better than benchmark models in terms of MAPEs. For 1 week in advance forecasting without weather features, stacked hybrid models achieve MAPEs of 12.28%, 13.09%, 13.16%, 12.63% for CNN-BiLSTM, CNN-LSTM, CNN-GRU and CNN-RNN respectively, while BiLSTM, LSTM, GRU, RNN, CNN and ConvLSTM models have MAPEs of 13.37%, 13.68%, 13.73%, 12.70%, 13.51% and 13.83% respectively. These hybrid models achieve better forecasting accuracies (up to 1.5%) than benchmark models in almost every case (RNN model is the only exception). With weather features, stacked hybrid models achieve MAPEs of 12.94%, 12.87%, 13.47%, 12.48% for CNN-BiLSTM, CNN-LSTM, CNN-GRU and CNN-RNN respectively, while BiLSTM, LSTM, GRU, RNN, CNN and ConvLSTM models have MAPEs of 13.02%, 13.85%, 14.74%, 12.73%, 13.28% and 14.77% respectively. Similarly, the hybrid models generally achieve better forecasting accuracies (up to 2.3%) than benchmark models. For 2 weeks in advance forecasting without weather features, stacked hybrid models achieve MAPEs of 15.39%, 17.20%, 16.99%, 16.72% for CNN-BiLSTM, CNN-LSTM, CNN-GRU and CNN-RNN respectively, while BiLSTM, LSTM, GRU, RNN, CNN and ConvLSTM models have MAPEs of 16.76%, 17.78%, 17.21%, 16.35%, 17.44% and 17.72% respectively, the hybrid models are still generally have better forecasting accuracies (up to 2.4%) than these benchmark models. For 2 weeks in advance forecasting with weather features, stacked hybrid models achieve MAPEs of 16.23%, 17.11%, 17.30%, 16.34% for CNN-BiLSTM, CNN-LSTM, CNN-GRU and CNN-RNN respectively, while BiLSTM, LSTM, GRU, RNN, CNN and ConvLSTM models have MAPEs of 16.78%, 18.03%, 18.13%, 16.74%, 17.83% and 18.08% respectively, the hybrid models are still generally have better forecasting accuracies (up to 1.9%) than these benchmark models. For 3 weeks in advance forecasting without weather features, stacked hybrid models achieve MAPEs of 18.51%, 20.89%, 20.74%, 20.84% for CNN-BiLSTM, CNN-LSTM, CNN-GRU and CNN-RNN respectively, while BiLSTM, LSTM, GRU, RNN, CNN and ConvLSTM models have MAPEs of 19.82%, 22.40%, 21.39%, 20.26%, 21.30% and 21.16% respectively, the hybrid models are still generally have better forecasting accuracies (up to 3.9%) than these benchmark models. For 3 weeks in advance forecasting with weather features, stacked hybrid models achieve MAPEs of 19.54%, 21.20%, 20.60%, 20.16% for CNN-BiLSTM, CNN-LSTM, CNN-GRU and CNN-RNN respectively, while BiLSTM, LSTM, GRU, RNN, CNN and ConvLSTM models have MAPEs of 19.73%, 22.35%, 22.04%, 20.55%, 21.59% and 21.88% respectively, the hybrid models are still generally have better forecasting accuracies (up to 2.8%) than these benchmark models. For 4 weeks in advance forecasting without weather features, stacked hybrid models achieve MAPEs of 21.72%, 24.25%, 24.69%, 24.93 for CNN-BiLSTM, CNN-LSTM, CNN-GRU and CNN-RNN respectively, while BiLSTM, LSTM, GRU, RNN, CNN and ConvLSTM models have MAPEs of 22.96%, 26.27%, 25.60%, 23.94%, 25.02% and 24.52% respectively. Some of the hybrid models have close forecasting accuracies to these benchmark models while the hybrid model CNN-BiLSTM still has better forecasting accuracies (up to 4.5%) than all these benchmark models. For 4 weeks in advance forecasting with weather features, stacked hybrid models achieve MAPEs of 22.90%, 25.15%, 23.49%, 23.74% for CNN-BiLSTM, CNN-LSTM, CNN-GRU and CNN-RNN respectively, while BiLSTM, LSTM, GRU, RNN, CNN and ConvLSTM models have MAPEs of 22.56%, 26.22%, 25.60%, 24.69%, 25.48% and 25.70% respectively, the hybrid models are still generally have better forecasting accuracies (up to 3.3%) than these benchmark models. From the above data, we can conclude that the stacked hybrid deep learning models have better performance, in terms of MAPEs, than their non-stacked counterparts in most of the cases with or without taking into account weather factors.

Another point to note is that with deep learning techniques and hybrid deep learning techniques, weather features such as temperature and rainfall do not contribute significantly to the overall performance in terms of MAPE. As can be seen from Figs. 3 and 4, the weather features only slightly improve the MAPEs of some of the deep learning models (non-hybrid and hybrid), while in most cases, the weather features have adversely effected the overall performance (in terms of MAPEs) of the deep learning models. This means that the deep learning models can learn the hidden patterns more effectively from the weekly dengue cases itself as compared to the correlation between the dengue cases and weather factors.

Furthermore, as demonstrated in this study, CNN-BiLSTM models have better performance (not only in terms of MAPEs but also in terms of MAEs and RMSEs) in almost all cases as compared to both the benchmark deep learning models (non-hybrid) and other hybrid deep learning models. This is likely because BiLSTM allows for additional training by traversing the input data from two directions, and can therefore better learn the hidden patterns in the data set as compared to the other recurrent networks.

In the proposed structure, as shown in Fig. 2, we mentioned that the structure can go deeper by adding more stacks of convolutional modules and recurrent modules. In the implementation stage, for each deep hybrid model, we have three versions: 1, 2 and 3 stacks. As we can see from Figs. 3 and 4, for different hybrid models, the performance is not necessarily improved with more hybrid stacks. For most cases, it appears that one stack of convolutional module and recurrent module is adequate for forecasting purposes. Specifically, the 1 stack CNN-BiLSTM model consistently outperforms the benchmark models and other hybrid deep learning models by 0.4–4.5% in terms of MAPEs. It also achieves better performance in terms of MAEs and RMSEs for almost all cases with and without consideration of the weather features.

Additionally, as shown in Figs. 3 and 4, the non-hybrid deep learning models also have good forecasting performance, proving the applicability of deep learning models in dengue cases forecasting.

As part of our future work, we intend to develop more hybrid deep learning models with other deep leaning techniques such as deep belief networks and the attention mechanism to further improve the forecasting performance. Forecasting daily dengue incidence in the region could also be another important research direction to pursue.

6. Conclusion

Accurate forecasting of dengue cases is very important because it allows for implementation of effective vector control measures, thereby reducing the healthcare and economic burden. Our developed deep learning hybrid models, comprising convolutional and recurrent layers, have proved to be effective in forecasting weekly dengue infections, as they can achieve good performance in terms of MAPE, MAE and RMSE for advance forecasting in terms of 1–4 weeks. The concept of hybrid architecture is novel in the context of forecasting dengue infection cases as we stack two types (hybrid) of deep learning models in dengue incidence forecasting. The best forecasting accuracy is 87.72% (1 week in advance), achieved by 1 stack CNN-BiLSTM model. This is the best weekly dengue incidence forecasting in terms of accuracy in the literature, outperforming the current best by 4.72%. The CNN-BiLSTM model has also consistently outperformed the non-hybrid benchmark models (RNN, GRU, LSTM, BiLSTM, CNN models) and other hybrid deep learning models by 0.4–4.5% (from marginal to significant improvement) in terms of MAPEs. Furthermore, the hybrid architecture does not need to be very deep, and it is noted that a deeper architecture does not necessarily translate to better forecasting performance. Generally, one stack of convolutional and recurrent modules is adequate for hybrid deep learning models. In our context, weather features such as temperature and rainfall do not contribute significantly to the overall forecasting performance as one will expect, and in contrary, have adverse effects on forecasting performance in most cases. The non-hybrid deep learning models used in this study also achieve very good forecasting performance, and are better than traditional machine learning models. The success as seen in applying deep hybrid models in forecasting dengue cases here could inspire researchers in the community to develop more deep learning based hybrid models for other applications involving time series forecasts.

CRediT authorship contribution statement

Xinxing Zhao: Software, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Kainan Li:** Formal analysis, Investigation, Writing – review & editing, Visualization. **Candice Ke En Ang:** Formal analysis, Investigation, Writing – review & editing, Visualization. **Kang Hao Cheong:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All data generated or analysed during this study are included in this published article.

Funding

This project was funded by the SingHealth-SUTD Population Health Innovation Fund (Grant No. PHIF 2021-02).

References

[1] Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, Drake JM, Brownstein JS, Hoen AG, Sankoh O, et al. The global distribution and burden of dengue. Nature 2013;496(7446):504–7.

[2] Cobra C, Rigau-Pérez JG, Kuno G, Vomdam V. Symptoms of dengue fever in relation to host immunologic response and virus serotype, Puerto Rico, 1990–1991. Am J Epidemiol 1995;142(11):1204–11.

[3] Guzmán MG, Kouri G. Dengue: an update. Lancet Infect Dis 2002;2(1):33–42.

[4] Nasar S, Rashid N, Iftikhar S. Dengue proteins with their role in pathogenesis, and strategies for developing an effective anti-dengue treatment: A review. J Med Virol 2020;92(8):941–55.

[5] Wilder-Smith A. Dengue vaccine development: status and future. In: Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz, vol. 63. Springer; 2020, p. 40–4, (1).

[6] Dorigatti I, McCormack C, Nedjati-Gilani G, Ferguson NM. Using Wolbachia for dengue control: insights from modelling. Trends Parasitol 2018;34(2):102–13.

[7] Martinez EZ, Silva EASd, Fabbro ALD. A SARIMA forecasting model to predict the number of cases of dengue in Campinas, State of São Paulo, Brazil. Rev Soc Bras Med Trop 2011;44:436–40.

[8] Withanage GP, Viswakula SD, Nilmini Silva Gunawardena Y, Hapugoda MD. A forecasting model for dengue incidence in the District of Gampaha, Sri Lanka. Parasites Vectors 2018;11(1):1–10.

[9] Siriyasatien P, Phumee A, Ongruk P, Jampachaisri K, Kesorn K. Analysis of significant factors for dengue fever incidence prediction. BMC Bioinformatics 2016;17(1):1–9.

[10] Scavuzzo JM, Trucco F, Espinosa M, Tauro CB, Abril M, Scavuzzo CM, Frery AC. Modeling Dengue vector population using remotely sensed data and machine learning. Acta Trop 2018;185:167–75.

[11] Xu J, Xu K, Li Z, Meng F, Tu T, Xu L, Liu Q. Forecast of dengue cases in 20 Chinese cities based on the deep learning method. Int J Environ Res Public Health 2020;17(2):453.

[12] Carvajal TM, Viacrusis KM, Hernandez LFT, Ho HT, Amalin DM, Watanabe K. Machine learning methods reveal the temporal pattern of dengue incidence using meteorological factors in metropolitan Manila, Philippines. BMC Infect Dis 2018;18(1):1–15.

[13] Liu Q, Liu X, Jiang B, Yang W. Forecasting incidence of hemorrhagic fever with renal syndrome in China using ARIMA model. BMC Infect Dis 2011;11(1):1–7.

[14] Kuan M-M, Lin T, Chuang J-H, Wu H-S. Epidemiological trends and the effect of airport fever screening on prevention of domestic dengue fever outbreaks in Taiwan, 1998–2007. Int J Infect Dis 2010;14(8):e693–7.

[15] Chakraborty T, Chattopadhyay S, Ghosh I. Forecasting dengue epidemics using a hybrid methodology. Phys A 2019;527:121266.

[16] Fan D, Sun H, Yao J, Zhang K, Yan X, Sun Z. Well production forecasting based on ARIMA-LSTM model considering manual operations. Energy 2021;220:119708.

[17] Harrou F, Dairi A, Kadri F, Sun Y. Forecasting emergency department overcrowding: A deep learning framework. Chaos Solitons Fractals 2020;139:110247.

[18] Livieris IE, Pintelas E, Pintelas P. A CNN–LSTM model for gold price time-series forecasting. Neural Comput Appl 2020;32(23):17351–60.

[19] Zhao X, Ang CKE, Acharya UR, Cheong KH. Application of Artificial Intelligence techniques for the detection of Alzheimer's disease using structural MRI images. Biocybern Biomed Eng 2021;41(2):456–73.

[20] Cheong KH, Tang KJW, Zhao X, Koh JEW, Faust O, Gururajan R, Ciaccio EJ, Rajinikanth V, Acharya UR. An automated skin melanoma detection system with melanoma-index based on entropy features. Biocybern Biomed Eng 2021;41(3):997–1012.

[21] Zhao X, Lai JW, Ho AFW, Liu N, Ong MEH, Cheong KH. Predicting hospital emergency department visits with deep learning approaches. Biocybern Biomed Eng 2022;42(3):1051–65.

[22] Yin X, Wu G, Wei J, Shen Y, Qi H, Yin B. Deep learning on traffic prediction: Methods, analysis and future directions. IEEE Trans Intell Transp Syst 2021.

[23] Somu N, MR GR, Ramamritham K. A deep learning framework for building energy consumption forecast. Renew Sustain Energy Rev 2021;137:110591.

[24] Law R, Li G, Fong DKC, Han X. Tourism demand forecasting: A deep learning approach. Ann Tour Res 2019;75:410–23.

[25] Shahid F, Zameer A, Muneeb M. Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. Chaos Solitons Fractals 2020;140:110212.

[26] Zhao X, Li K, Ang CKE, Ho AFW, Liu N, Ong MEH, Cheong KH. A deep learning architecture for forecasting daily emergency department visits with acuity levels. Chaos Solitons Fractals 2022;165:112777.

[27] Lv Y, Duan Y, Kang W, Li Z, Wang F-Y. Traffic flow prediction with big data: a deep learning approach. IEEE Trans Intell Transp Syst 2014;16(2):865–73.

[28] Zhang Z, Jing Q, Chen Z, Li T, Jiang L, Li Y, Luo L, Marshall J, Yang Z. The increasing menace of dengue in Guangzhou, 2001–2016: the most important epicenter in mainland China. BMC Infect Dis 2019;19(1):1–8.

[29] Mussumeci E, Coelho FC. Large-scale multivariate forecasting models for Dengue-LSTM versus random forest regression. Spatial Spatio-Tempor Epidemiol 2020;35:100372.

[30] Nguyen V-H, Tuyet-Hanh TT, Mulhall J, Minh HV, Duong TQ, Chien NV, Nhung NTT, Lan VH, Minh HB, Cuong D, et al. Deep learning models for forecasting dengue fever based on climate data in Vietnam. PLoS Negl Trop Dis 2022;16(6):e0010509.

[31] Sanchez-Gendriz I, de Souza GF, de Andrade IG, Neto ADD, de Medeiros Tavares A, Barros D, de Morais AHF, Galvão-Lima LJ, de Medeiros Valentim RA. Data-driven computational intelligence applied to dengue outbreak forecasting: a case study at the scale of the city of Natal, RN-Brazil. Sci Rep 2022;12(1):1–10.

[32] Saleh AY, Baiwei L. Dengue prediction using deep learning with long short-term memory. In: 2021 1st International conference on emerging smart technologies and applications (ESmarTA). IEEE; 2021, p. 1–5.

[33] Shi Y, Liu X, Kok S-Y, Rajarethinam J, Liang S, Yap G, Chong C-S, Lee K-S, Tan SS, Chin CKY, et al. Three-month real-time dengue forecast models: an early warning system for outbreak alerts and policy decision support in Singapore. Environ Health Perspect 2016;124(9):1369–75.

[34] Chimmula VKR, Zhang L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. Chaos Solitons Fractals 2020;135:109864.

[35] Chen A, Wang F, Liu W, Chang S, Wang H, He J, Huang Q. Multi-information fusion neural networks for arrhythmia automatic detection. Comput Methods Programs Biomed 2020;193:105479.

[36] Yan T, Wong PK, Ren H, Wang H, Wang J, Li Y. Automatic distinction between COVID-19 and common pneumonia using multi-scale convolutional neural network on chest CT scans. Chaos Solitons Fractals 2020;140:110153.

[37] Hassantabar S, Ahmadi M, Sharifi A. Diagnosis and detection of infected tissue of COVID-19 patients based on lung X-ray image using convolutional neural network approaches. Chaos Solitons Fractals 2020;140:110170.

[38] Swapnarekha H, Behera HS, Nayak J, Naik B. Role of intelligent computing in COVID-19 prognosis: A state-of-the-art review. Chaos Solitons Fractals 2020;138:109947.

[39] Abbasimehr H, Paki R. Prediction of COVID-19 confirmed cases combining deep learning methods and Bayesian optimization. Chaos Solitons Fractals 2021;142:110511.

[40] Mehtab S, Sen J. Analysis and forecasting of financial time series using CNN and LSTM-based deep learning models. In: Advances in distributed computing and machine learning. Springer; 2022, p. 405–23.

[41] Salim NAM, Wah YB, Reeves C, Smith M, Yaacob WFW, Mudin RN, Dapari R, Sapri NNFF, Haque U. Prediction of dengue outbreak in Selangor Malaysia using machine learning techniques. Sci Rep 2021;11(1):1–9.

[42] Sousa LB, Zuben CJV, Zuben FJV. Regularized linear and nonlinear autoregressive models for dengue confirmed-cases prediction. In: International conference on unconventional computation and natural computation. Springer; 2015, p. 121–31.