Chaos, Solitons and Fractals 165 (2022) 112777

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

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

A deep learning architecture for forecasting daily emergency department visits with acuity levels

Xinxing Zhao [a], Kainan Li [a], Candice Ke En Ang [b], Andrew Fu Wah Ho [c,d], Nan Liu [e], Marcus Eng Hock Ong [c,e], 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
[c] Department of Emergency Medicine, Singapore General Hospital, Singapore
[d] Pre-hospital and Emergency Research Center, Duke-NUS Medical School, Singapore
[e] Health Services and Systems Research, Duke-NUS Medical School, Singapore

ARTICLE INFO

Keywords:
Deep neural networks
Emergency department
Patient visit forecasting
Healthcare
Time series forecasting

ABSTRACT

Accurate forecasting of Emergency Department (ED) visits is important for decision-making purposes in hospitals. It helps to form tactical and operational level plans, which facilitates staff and resource allocations in advance. A dataset recording the daily visits of patients at the ED of a regional hospital over a 3-year period is used in this study. Patients are triaged into 3 acuity levels: P1, P2 and P3, with P1 being patients with severe or life threatening conditions, whereas P3 being patients with minor injuries requiring less urgent attention. A novel deep learning forecasting structure, which has advantages of both Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN), is being developed and applied to forecasting daily visits (up to 56 days into the future) for the different acuity levels. The features included in this study are calendar days, public holidays, Pollution Standard Index (PSI) readings, rainfall and daily average temperature. The effectiveness of our newly developed model, in terms of forecasting accuracy, is demonstrated and compared with other deep learning models. Our model achieves mean absolute percentage errors (MAPEs) of 17.37%, 7.19%, 6.11% and 4.50% in forecasting P1, P2, P3 and total visits respectively, and has demonstrated superior performance when evaluated against state-of-the-art studies in the literature. This study illustrates that utilization of our hybrid model comprising LSTM with CNN layers can provide a significant improvement over these existing deep learning models for ED daily visits forecasting.

1. Introduction

Forecasting of visits to the Emergency Department (ED) is important for the decision-making of hospitals at different levels of management, as it helps to form strategic, tactical and operational plans. For instance, such may include capacity planning in the long term, whereby emergency facilities expansion are planned months or years ahead, or short term plans involving staff and resource allocation days to weeks ahead.

As the demand for healthcare rapidly increases throughout the world [1–4], overcrowding in EDs is a resulting issue that has been and will remain for healthcare systems. Overcrowding causes long waiting times [5], higher dissatisfaction rates [6], increased risk of mortality [7], and morale and productivity issues [8]. Improving ED capabilities often require more staff and resources. However, by forecasting accurately the timing and scale of ED visits [9,10], staff and resources can be allocated appropriately with aims to meet the demand, and therefore alleviating the issues faced with overcrowded EDs. There are mainly two categories [11] of models for the time series forecasting problem: statistical and machine learning models. Traditional statistical models such as ARIMA [12,13] are mainly linear and may not model the stochastic and non-linear nature of time series data well, while more advanced machine learning models [14] such as artificial neural networks [15] and support vector machine [16] are more adapted to these challenges. Deep learning models use layered architectures to extract different features in data from different layers and have shown stronger capabilities [17,18] in solving many time-series prediction problems as compared to more traditional statistical and machine learning models. For example, more advanced models have successfully been applied in many domains such as traffic [19], stock market [20], oil production [21], tourism [22] and COVID-19 [23] forecasting. There are also studies that use hybrid models which combine different models

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

https://doi.org/10.1016/j.chaos.2022.112777
Received 18 August 2022; Received in revised form 29 September 2022; Accepted 2 October 2022
Available online 21 October 2022
0960-0779/

Table 1
Recent studies involving ED patient arrivals forecasting with deep learning methods.

Reference	Data length& Used variables	Methods & Horizon	Performance (MAPE or R^2)
This study (2022)	3 years; Calendar, public holidays, temperature, rainfall and air pollution measured by PSI readings	Proposed a stacked architecture based on LSTM and CNN; Daily	MAPE of 4.50% for 1 day, 4.62% for 3 days in advance, 4.71% for 7 days in advance and 5.02% for 56 days in advance forecasting.
[29] (2022)	396 days; Meteorological, calendar and constructed features	LSTM and other machine learning techniques, using kernel principal component analysis and maximal information coefficient method to deal with the features; Daily and hourly	MAPE of 40%–70% hourly prediction, daily 9%–13%.
[30] (2022)	5 years, 3 years, 1 year and 6 months; day of week	Stacked architecture with LSTM, RNN and GRU as components; Daily	MAPE of prediction, daily 5.41–5.72%.
[18] (2021)	3.5 years; Meteorological and calendar	Proposed new structures for CNN, LSTM and RF models; Daily and weekly	MAPE of current day LSTM 8.04%, 3 days in advance CNN 9.24%, 7 days in advance LSTM 8.91%.
[31] (2020)	13 years; Meteorological, pollen and chemical pollution and hospital calendar	Proposed a new deep learning structure based on CNN and LSTM; Daily	R^2 of 0.93 and 0.90 for circulatory and respiratory cases, respectively.
[32] (2020)	3 years; Calendar	LSTM and RNN; Daily and hourly	Daily R^2 of 0.967, 0.972, hourly R^2 of 0.918, 0.914 for LSTM and GRU models.
[11] (2020)	3 years; Day of week, month of year	Variational Autoencoders for one-step and multi-steps forecasting and LSTM, RNN, GRU, CNN; Hourly and daily	Hourly and daily R^2 of 0.949, 0.925, respectively.

together for time series forecasting problems. For example, a traditional statistical model ARIMA and a wavelet transform-based model are combined together for COVID-19 cases [24] forecasting, a LSTM model combined with empirical wavelet transform is used for digital currency [25] forecasting, while CNN and LSTM models are combined for gold price forecasting [26]. These studies have demonstrated that more advanced models and better forecasting performance can be achieved by hybridizing different models. There are also other models, such as iteration models [27] or nature-inspired algorithms [28] that can be used for time series forecasting problems.

In this study, we focus on deep learning models and discuss a novel hybrid deep learning architecture possessing advantages of both CNN and LSTM models that has been developed to forecast daily ED visits. The forecasting performance of this architecture is compared against leading deep learning models for time series forecasting such as RNN, LSTM, biLSTM, ConvLSTM, GRU, CNN and also with state-of-art models in the literature. The results indicate that the developed hybrid models are more accurate than existing deep learning methods in forecasting daily ED visits.

The main contributions of this study are as follows: firstly, the work presented here is one of the first studies that uses a hybrid stacked architecture with CNN and LSTM models as components in the daily ED visits forecasting. The demonstrated effectiveness in terms of forecasting accuracy is superior to the state-of-the-art in the literature. This could lead to the development of different combinations of network typologies and more hybrid deep learning models that could improve forecasting performance in other domains as well, and not only limited to ED visits. Secondly, to better understand the ED patient dynamics in terms of timing and scale and for better planning, the forecast horizon is up to 56 days into the future, which is sufficient to improve resource allocation in advance. Thirdly, in contrast to most existing studies, by taking account of forecasting ED visits with different acuity categories (P1, P2 and P3 visits), it provides a valuable way for micro level planning, as the resources required for patients will differ in terms of their urgency for medical attention. Lastly, the deep learning models, including our newly developed models can be retrained easily and therefore they can be used as a quasi-real-time forecasting tool for online platforms.

The rest of this paper is organized as follows: Section 2 discusses the related work in this topic, whereas Section 3 provides details of the dataset utilized, the proposed architecture, as well as the evaluation metrics and configurations of the deep models. Section 4 and Section 5 provide the results, discussion and analysis of the models. Finally, our conclusion is presented in Section 6.

2. Related work

Emergency Departments face huge volumes of patient visits daily [18,33], and by forecasting patient visits timely and accurately, patients may be managed more effectively, and therefore the issues related to overcrowding in the ED can be alleviated. Many studies use different time series models in forecasting ED patient visits are available in literature. Sudarshan et al. [18] and Zhao et al. [30] have summarized some of these studies and models.

Apart from statistical models and traditional machine learning models, more recently, deep recurrent neural networks and its variants, such as LSTM and GRU [34] are frequently used in time series forecasting problems, as they have the ability to store and manipulate long sequences of data. Other deep learning models such as CNN and variational autoencoder are also adapted or modified in forecasting problems. In a French study, Harrou et al. [11] proposed a deep learning method that uses variational autoencoder algorithm to forecast daily and hourly visits at an ED. The autoencoder algorithm has a better or comparable performance to the leading deep learning models such as LSTM, GRU, RNN and CNN, and reached R^2 of 0.949 and 0.925 for hourly and daily respectively. The study also demonstrated that the proposed method has better performance than the benchmark models with multi-steps forecasting (forecasting more days and hours

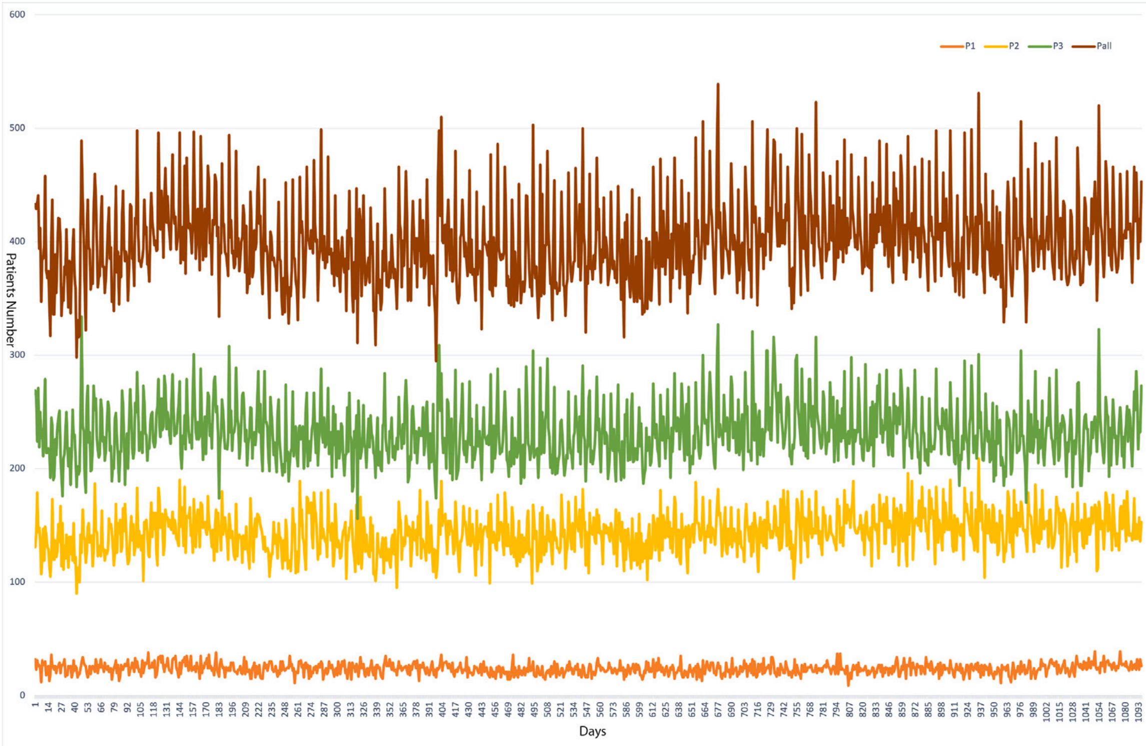

Fig. 1. Daily visits at SGH ED by patient acuity levels.

into future). Zhang et al. [29] used ten machine learning methods, including LSTM, with calendar and meteorological information as well as two feature selection methods to forecast patient arrivals. Their models achieved MAPEs of 45%–70% and 8.8%–13% for hourly and daily predictions, and concluded that calendar and meteorological information have significant influence on the predictions. Zhao et al. [30] used a stacked deep learning architecture, with RNN, LSTM and GRU as its components to predict daily ED visits. One important aspect of this study is that they tested their different deep stacked models with different lengths of dataset (5 years to 6 months). They concluded that for different lengths of dataset, different number of hidden layers and hidden units should be used to achieve the best performance. The MAPEs of their daily ED prediction are in the range 5.41–5.72%. Sudarshan et al. [18] proposed their own versions of RF, LSTM and CNN models for ED visits forecasting, they considered 1 day, 3 days and 7 days into the future with weather information. For 1 day into the future forecasting, LSTM achieved a MAPE of 8.04%, CNN 9.53% and RF model 10.10%. For 3 days into the future forecasting, the CNN model reached an average MAPE of 9.24% while for 7 days forecasting, LSTM reached an average MAPE of 8.91%. In their study, deep neural networks based models have better performance than these RF models in all 1 day, 3 days and 7 days forecasting cases. Navares et al. [31] used hybrid deep learning models to forecast daily hospital admissions due to respiratory and circulatory problems and compared their performance with ARIMA, RF, GBM and ANN models. The proposed hybrid architecture which combined LSTM and CNN achieved an average R^2 of 0.93 and 0.90 for the circulatory and respiratory cases, outperforming the statistical and more traditional machine learning models in almost all instances.

As we can see from Table 1, researchers use different forecasting variables, such as calendar and meteorological information, as features to boost forecasting performance of their models. As for the horizons for forecasting, there are hourly (current hour or a few hours into the future), daily (1 day to 7 days into the future) and weekly forecasts, which are well suited for operational and tactical planning for the hospitals. We will discuss more about the methods, forecasting horizons and performance of the models in the discussion part.

Table 2
Descriptive statistics for the 3-year dataset (daily visits) based on acuity categories.

Category	Max	Min	Mean	Median	SD
P1	39	9	24.32	24	4.77
P2	209	89	142.35	143	18.03
P3	358	156	229.13	229	25.63
Pall	539	295	396.26	395	37.71

3. Methodology

In this section, we first introduce the 3-year ED visits dataset, and provide details about the proposed stacked hybrid architecture that is used for forecasting of daily visits. We then give a short introduction to the deep learning models used in this study and their configuration details. Finally, we discuss the evaluation metrics used to compare different deep learning models.

3.1. The dataset

Singapore General Hospital (SGH) is the first and largest hospital in Singapore. The ED at SGH faced an average demand of 396 patients (maximum 539, minimum 295, median 395 and SD 37.71) daily in the year of 2010 to 2012. As seen from Fig. 1, daily visits are categorized into 3 different emergency acuity levels [35], namely P1, P2 and P3. P1 cases consist critically ill patients whereby immediate treatment is needed, P2 being acutely ill patients, P3 being less acutely ill patients who can endure some delay to receive treatment. Apart from these 3 acuity levels, the total daily visits to the ED (Pall) are also being considered in this study. More statistical details about these categories can be found in Table 2.

As we want to forecast the visits of each category, P1, P2, P3 and the total ED (Pall) visits will be divided separately into a training set and a testing set (chronologically ordered) in the development of our deep learning models. The ratio is 7:3 (training set vs testing set), similar to other deep learning based forecasting studies [11,19].

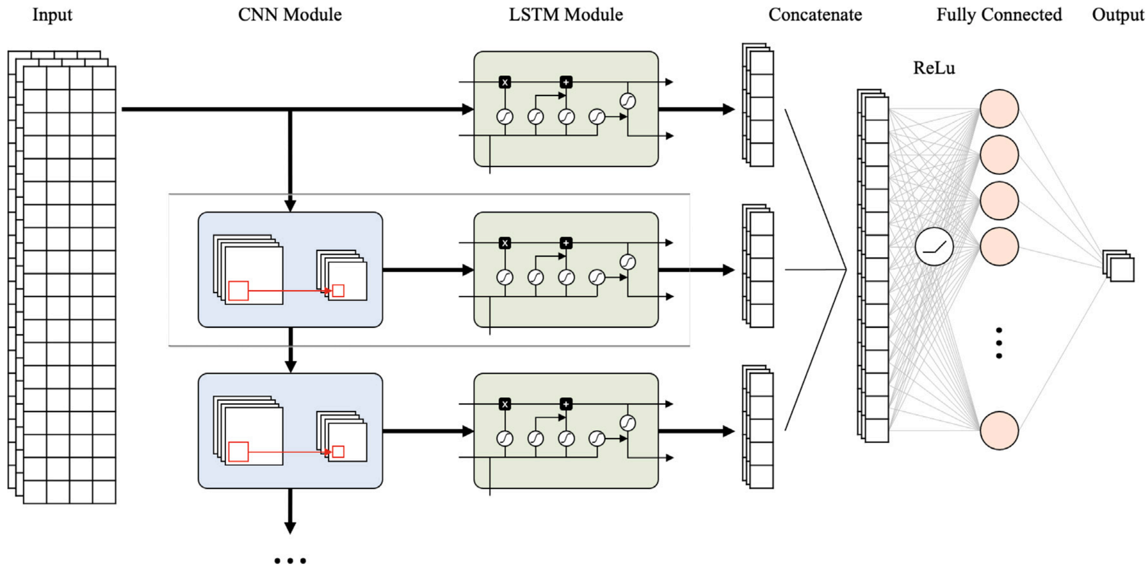

Fig. 2. The proposed structure for daily visits forecasting. One module of CNN and one module of LSTM is a hybrid layer. This architecture can go deeper by adding more hybrid layers.

3.2. The proposed architecture

Long short-term memory networks (LSTM) [36,37] are a variation of recurrent neural networks (RNNs). They use gate structures (input, output and forget gate) to control the information flow and are capable of learning long-term dependencies with cell states. LSTM networks can remember previous network information and connect it with the present ones, hence they have better capabilities over more traditional recurrent networks on temporal forecasting tasks [37,38].

CNN networks have been widely applied to image related applications, such as image classification and analysis [39,40], object detection [41]. However, CNN networks are not limited to dealing with images and can be used in time series forecasting [42,43] as well. CNN networks are characterized by their abilities to automatically detect important features and can be used to learn internal structures of time-series data, while LSTM networks are good at identifying short-term dependencies and subsequently using these short-term dependencies to build long-term dependencies. Therefore, by considering and combining the features of these two deep learning techniques, a structure which has advantages of both LSTM and CNN could boost performance for forecasting ED visits.

With this in mind, we have proposed a deep structure, as shown in Fig. 2, named stacked LSTM-CNN, which has two components: CNN and LSTM layers. CNN layers are used for feature extraction from data input and LSTM layers are used to predict the sequence of visits. As can be seen, the input signals go through the first LSTM module directly, and at same time the first CNN module then an LSTM module (a hybrid layer consists these two modules combined). The signals can go through more hybrid layers and the structure can go deeper. The processed signals output from LSTM modules will be concatenated, go through a fully connected layer with ReLu, then again through another fully connected layer before the output. Notice that the input signals go to the first LSTM module directly in the structure, allowing LSTM to work independently to deal with the time series sequence and this could help to improve the overall performance of the forecasts.

The forecasting horizons in this study are: 1, 3, 7 days and 56 days into the future. These forecasting horizons are important as they allow hospitals to form and adjust both their operational and tactical plans accordingly. It is worth noting that the P1, P2, P3 and total visits will be forecast separately, as this will likely help the hospital management team with planning both at the micro and macro levels.

3.3. Models and configurations and features

As mentioned earlier in the structure as shown in Fig. 2, the signals can go through more hybrid layers and the structure can go deeper. Variations of models can be derived by changing the number of hybrid layers (a hybrid layer has a CNN module and a LSTM module). We call the model with 1 layer of CNN and LSTM modules the version 1 model (V1), 2 layers the version 2 model (V2), 3 and 4 layers the version 3 and 4 model (V3 and V4). The version 5 model (V5) is given as shown in Fig. 3, it has 2 hybrid layers (instead of 5) of modules, and the first LSTM module (as circled) is removed from the structure. Hence, the main difference between V5 and V2 is the exclusion of this first LSTM module, while V1 to V4 models all have this first LSTM module.

The CNN module in our proposed structure has a convolutional 2D part with kernel size (3, 1) and padding (1, 0). The kernel deals with the features while the padding is used to maintain the shape. For the LSTM module in the proposed structure, it has 3 layers, each layer has 50 hidden units and drop rate of 0.25.

We compare our stacked hybrid models with other deep learning models such as LSTM [44], biLSTM [45], ConvLSTM [46], RNN [47], GRU [48] and CNN [26]. More details about the different structures and how the signals are processed in these deep learning models can be found in the above papers. For the configurations of the deep learning models in this study, they are as follows: for RNN, GRU, LSTM and biLSTM, they have same settings as the LSTM module in the proposed structure, each of them has 3 layers, each layer has 50 hidden units and drop rate of 0.25, and ReLu as the activation function. For CNN, it has 3 convolutional layers, each layer with kernel size 2 and stride 2, drop rate of 0.25, ReLu as the activation function. For CovLSTM, it has two convolutional layers with kernel size 3, padding 1, 2 max pooling layers with kernel size 2 and stride 2, while LSTM part has 3 layers, each layer has 50 hidden units and drop rate of 0.25, ReLu as the activation function.

The features used in this study are calendar days (Monday effect considered), public holidays, PSI readings, rainfall and daily average temperature.

3.4. Evaluation metrics

To assess the performance of our forecasting models for the daily ED visits, the Mean Absolute Percentage Error (MAPE) is used. In Eq. (1),

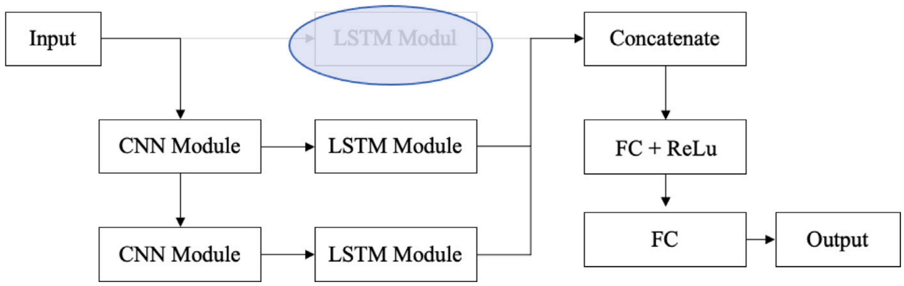

Fig. 3. Variations (V5) of proposed structure for daily visits forecasting.

Label	Day	Stacked CNN-LSTM Model					Benchmark Models					
		V1	V2	V3	V4	V5	CNN	RNN	GRU	LSTM	ConvLSTM	biLSTM
P1	1	0.175066	0.177539	0.173740	0.177387	0.177564	0.191266	0.177205	0.176921	0.175204	0.177564	0.176027
	3	0.176879	0.178817	0.176034	0.178687	0.178836	0.191863	0.178483	0.179231	0.177619	0.178836	0.178321
	7	0.178298	0.179050	0.178163	0.178917	0.179071	0.195932	0.178792	0.181479	0.178546	0.179072	0.178848
	56	0.173298	0.173641	0.173658	0.173498	0.173654	0.186624	0.173326	0.173337	0.173586	0.173654	0.173199
P2	1	0.071895	0.072481	0.073333	0.072077	0.071993	0.081290	0.088783	0.084426	0.084966	0.094996	0.088415
	3	0.072671	0.073394	0.074528	0.073322	0.073152	0.082625	0.089503	0.086381	0.086313	0.095959	0.090445
	7	0.073592	0.074607	0.075151	0.074884	0.074046	0.083000	0.090140	0.087021	0.086357	0.097204	0.092272
	56	0.075117	0.076188	0.076429	0.077308	0.075182	0.081491	0.091973	0.085428	0.088771	0.099522	0.095764
P3	1	0.062393	0.062279	0.063388	0.062146	0.061148	0.064473	0.076730	0.065545	0.064386	0.069066	0.070250
	3	0.062237	0.063601	0.063425	0.062458	0.061766	0.066406	0.077261	0.065939	0.064285	0.069989	0.073171
	7	0.062004	0.065764	0.064238	0.064965	0.062875	0.066759	0.076554	0.065039	0.064481	0.069184	0.074119
	56	0.068717	0.072154	0.070301	0.072355	0.068396	0.070070	0.079787	0.070169	0.068718	0.071625	0.080472
Pall	1	0.045594	0.045381	0.045214	0.045011	0.046523	0.049344	0.062557	0.047375	0.049219	0.053238	0.046817
	3	0.046611	0.046760	0.047342	0.046255	0.047438	0.050362	0.063405	0.048205	0.049983	0.054022	0.047931
	7	0.047118	0.047534	0.049384	0.047261	0.048162	0.050843	0.063586	0.048665	0.050293	0.053991	0.048467
	56	0.050189	0.050464	0.053777	0.050690	0.050860	0.052841	0.066417	0.051394	0.053079	0.056363	0.051681

Fig. 4. Deep learning models performance for 1, 3, 7, 56 days in advance forecasting in terms of MAPEs. Generally the 1 hybrid layer V1 model has the best performance in all cases. And it achieves about 95.5% for all ED visits for 1 day in advance forecasting. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

n represents number of data, y_t is the observed number at day t, and $\hat{y}_t$ is the forecasted value at day t.

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

4. Results

Fig. 4 gives the performance of different deep learning models in terms of MAPEs. The labels are for the different categories (P1, P2, P3 and Pall), and days describe the horizons of the models that forecast into the future (1, 3, 7 and 56 days). In each row from the third, green marked figures display better performance in terms of MAPE (the lower the MAPE, the better the performance, the greener the figure), yellow are intermediate while the red displays the worst performance for a certain label and horizon. For example, in the third row, 0.173740 (the smallest number in that row) is the best MAPE provided by V3 model that forecast P1 patients for 1 day. As we can see for all categories, the newly developed deep learning models have largely better or equivalent performance when compared with other benchmark deep learning models. The developed deep learning models achieved an accuracy of more than 95% (1-MAPE) for total daily patients (Pall), about 94% for P3, about 93% for P2 and a bit more than 82% for P1 visit forecasts.

For the P1 category, the developed deep learning models have small margin over these benchmark deep learning models for the different horizons except for the CNN models, whereby our models outperform CNN models about 1%–2%.

For the P2 category, the developed models have better performance than these benchmark deep learning models across all horizons. For each horizon, the developed models are about 1%–2% more accurate. For the P3 category, the best developed model is V5 and it has 0.2–1.6% better performance than the benchmark models for the forecasts of all horizons. For the Pall category, the best developed models have 0.1–1.6% better performance than these benchmark models for the forecasts of all horizons.

For each category (P1, P2, P3 and Pall) across all deep learning models, the forecasting performance generally decreases as the forecast duration extends. For example, V1 model has MAPEs of 0.0456, 0.0466, 0.0471 and 0.0502 for 1 day, 3 days, 7 days and 56 days forecasts into the future for the Pall category. However, for the P1 category, for the forecast of 56 days, all the deep models have slight better performance than 1 day, 3 days and 7 days forecasts. This is because majority of P1 visits are less predictable and likely involve more variables such as human and disease factors on top current variables including calendar and environmental factors.

Fig. 5 gives the best performance for P1, P2, P3 and Pall forecasting (1 day in advance) of our developed models. The used models for getting the best performance are V3, V1, V5, V4 for P1, P2, P3 and Pall forecasting, respectively. Fig. 6 gives the scatter plots of actual versus forecasted with these models for 1 day in advance forecasting. As noted from these two figures, the developed models have good performance in forecasting for 1 day in advance, except for the P1 category. As Fig. 4 has already demonstrated, the developed models have slightly decreasing performance as duration of forecast extends from 1, 3, 7 to 56 days for all the models except for P1 at 56 days forecasting. We also note that in each of these forecasting horizons, the developed models have 0.1%–2% more accurate performance than these benchmark deep learning models. Therefore, we will not include

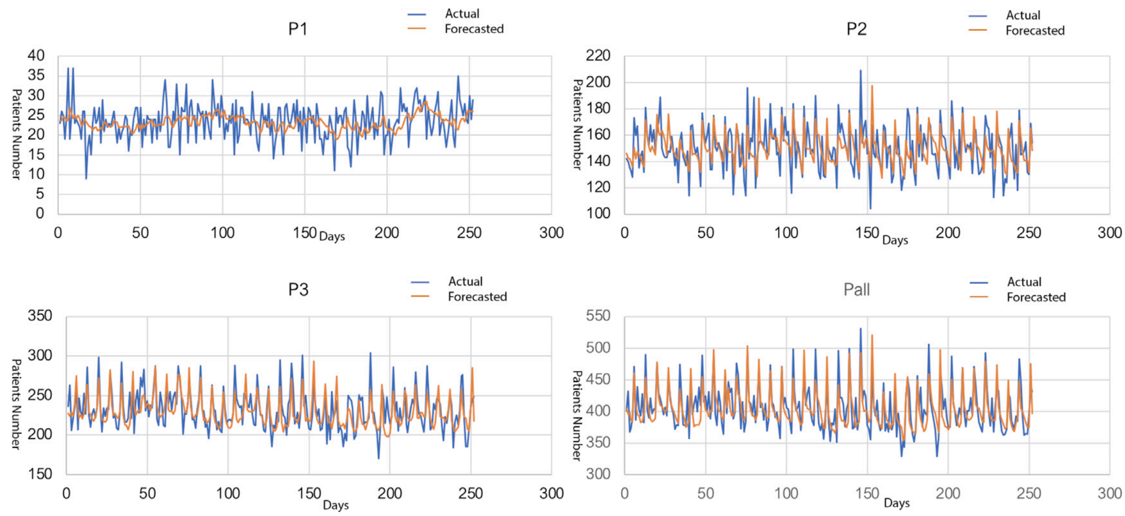

Fig. 5. Forecasting of daily visits for P1, P2, P3 and Pall. The models achieve very good performance for P2, P3 and Pall categories and reasonable performance for P1.

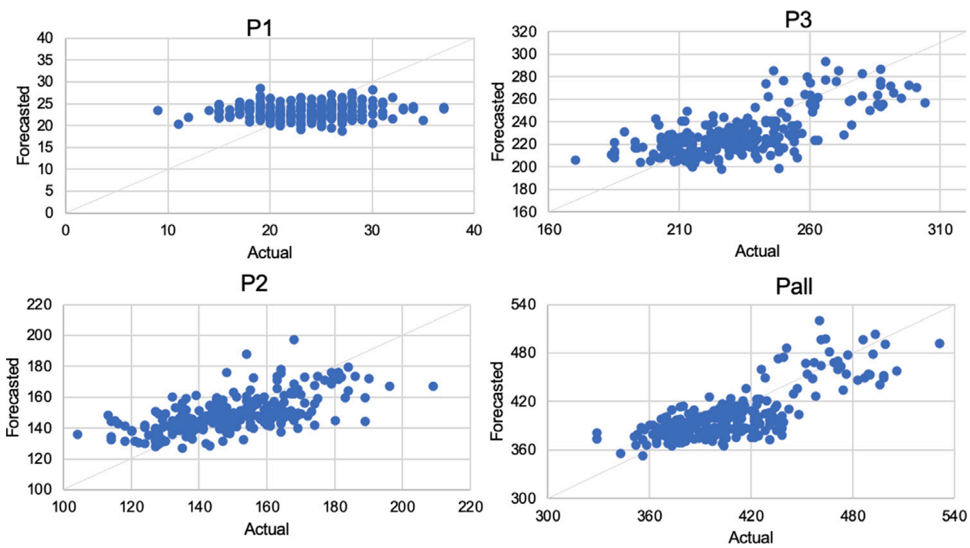

Fig. 6. The scatter plots of actual versus forecasted patient visits at the ED for each category (best performance).

more performance figures that with different horizons and categories for our developed and benchmark models.

5. Discussion

As using deep learning techniques for forecast of ED visits is relatively new, we have summarized the recent studies in Table 1. RNN and its variants such as LSTM and GRU are the most frequently used deep learning models for forecasting ED visits in the literature, whereas other deep learning techniques such as CNN and VAE are also adapted in this area. In this study, we use a hybrid structure which stacks LSTM and CNN models for forecasting ED visits. As demonstrated (Fig. 4), the developed models can achieve good performance, and are consistently better than the benchmark models. As far as we know, about 95.5% accuracy (1-MAPE) for the total (Pall) ED daily visits (with horizons of 1, 3, 7 and 56 days into future) forecasting displayed the best result among all the studies in the literature which use machine learning models or other traditional statistical methods. Our study also takes into consideration the different acuity levels. The developed models achieved good accuracy for P2 (about 93%) and P3 (about 94%) visits, and moderate accuracy for P1 (about 82%) visits forecasting. For the P1 patients, during the study period, the daily number of visits ranged from 9 to 39. The small size of this category could contribute to the difficulty of forecasting. However, the developed models still demonstrated relatively good forecasting ability. The developed models outperform both benchmark models (as demonstrated in Fig. 4) and other well established research summarized in [18] for daily forecasting in terms of MAPE. As compared to the most recent deep learning based research [18,29], for the total ED daily visits (Pall) with 1 day in advance forecasting, the developed models have 3.5–8.5% more accurate performance, while with 3 days and 7 days in advance forecasting, the developed models have 4.2–4.6% more accurate performance. Our developed models also demonstrate very strong capabilities for 56 days in advance forecasting, whereas most of the studies available in literature do not support such long-term forecasting.

For the forecasting of different duration into the future, there are hourly, daily, weekly and monthly studies available in the literature. In this study, we focused on daily (1 day, 3 days, 7 days and 56 days) forecasting into the future, as this could help the hospital management team to form short-term and long-term plans. However, hourly forecasting could help to form more refined plans at the operational level, and this could be one of our future research directions.

It is worth noting that, as can be seen from Fig. 4, the forecasting performance of the developed models do not change much when more hybrid layers of LSTM and CNN modules are added to the architecture. Some studies [49,50] have demonstrated that by increasing the layers of deep learning models, the performance can be improved. This is true only to a certain level, for example in [11,19], where 2–5 layers are used for different scenarios and for different deep learning models to achieve their best performance. In this study, we set each LSTM and CNN module in the structure to have 3 layers and we want to test similarly whether more layers of hybrid modules will increase the overall forecasting performance. As explained, the differences among developed models V1, V2, V3, and V4 are only in the hybrid layers of modules. By adding hybrid layers of modules to V1, we get other versions of V2, V3 and V4. As can be seen, V1 in most cases have a slightly better performance as compared to V2–V4. This means that the V1 model (V1 model has one layer of hybrid modules plus a separate LSTM module) is already adequate for our forecasting purpose. Another point to note is that V2 model has comparable performance with V5 model in most of the cases; this means a LSTM module difference between these two versions does not effect the forecasting performance much.

It is worth noting that our benchmark models have better performance than the most recent studies as shown in Table 1. This is due more to the features that are being utilized than to the optimal configurations used for the benchmark deep learning models. The features included are day of the week, week of the month, month of the year, public holidays, daily ambient average temperature and rainfall and PSI readings. These features take into account many internal patterns of the dataset, such as the Monday and holiday effect [51,52], weekly and monthly variations [53]. Temperature and rainfall are also used, as there are studies [18,29] that have demonstrated improved forecasting accuracy of ED visits by incorporating meteorological information as features. Notably, high PSI readings is linked to increased ED visits and hospital admissions [54] for respiratory and cardiac related conditions [55].

As part of our future work, we can also utilize other deep learning hybrid models such as RNN+CNN, BiLSTM+CNN, GRU+CNN. Attention mechanism can also be applied to the forecasting of Emergency Department visits.

6. Conclusion

Timely and accurate forecasting of daily ED visits is of utmost importance in hospital planning. Forecasting of visits in advance can help the management allocate resources more efficiently in both operational and tactical levels and can help to alleviate the issues related to ED overcrowding, thereby providing more satisfactory services to patients. In this work, an original and novel deep learning architecture, known as LSTM-CNN, is being proposed for the forecasting of patient visits at the ED. We have considered ED visits categorized into the different acuity levels in addition to the total daily visits, as this will help with planning of hospital resources at both the micro- and macro-levels. Our study has also demonstrated that although RNN models and their variants such as LSTM and GRU models are more widely used for time series forecasting studies in the literature, their utilization along with additional convolutional layers can provide a significant boost (in many cases, the forecasting accuracy improvement could reach 2%) in increasing forecasting performance. In our case, we stacked LSTM with CNN models, and the resulting structure demonstrated better performance than either LSTM or CNN models alone, as well as similar research with deep learning techniques in the literature. The effectiveness of this novel deep learning structure in terms of forecasting accuracy can certainly lead to development of more deep learning hybrid models to further improve the forecasting performance.

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. **Andrew Fu Wah Ho:** Methodology, Formal analysis, Writing – review & editing, Visualization. **Nan Liu:** Methodology, Formal analysis, Writing – review & editing, Visualization. **Marcus Eng Hock Ong:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition. **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.

Funding

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

Data availability

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

References

[1] Aboagye-Sarfo P, Mai Q, Sanfilippo FM, Preen DB, Stewart LM, Fatovich DM. A comparison of multivariate and univariate time series approaches to modelling and forecasting emergency department demand in western Australia. J Biomed Inform 2015;57:62–73.
[2] Chia NC, Loh SP. Using the stochastic health state function to forecast healthcare demand and healthcare financing: Evidence from Singapore. Rev Dev Econ 2018;22(3):1081–104.
[3] Yang P, Yang G, Qi J, Sheng B, Yang Y, Zhang S, Bi G, Mao X. The effect of multiple interventions to balance healthcare demand for controlling COVID-19 outbreaks: a modelling study. Sci Rep 2021;11(1):1–13.
[4] Vollmer MA, Glampson B, Mellan T, Mishra S, Mercuri L, Costello C, Klaber R, Cooke G, Flaxman S, Bhatt S. A unified machine learning approach to time series forecasting applied to demand at emergency departments. BMC Emerg Med 2021;21(1):1–14.
[5] Khanna S, Boyle J, Good N, Bell A, Lind J. Analysing the emergency department patient journey: discovery of bottlenecks to emergency department patient flow. Emerg Med Aust 2017;29(1):18–23.
[6] Abidova A, da Silva PA, Moreira S. Predictors of patient satisfaction and the perceived quality of healthcare in an emergency department in Portugal. West J Emerg Med 2020;21(2):391.
[7] Jo S, Jeong T, Jin YH, Lee JB, Yoon J, Park B. ED crowding is associated with inpatient mortality among critically ill patients admitted via the ED: post hoc analysis from a retrospective study. Am J Emerg Med 2015;33(12):1725–31.
[8] Oh HC, Chow WL, Gao Y, Tiah L, Goh SH, Mohan T. Factors associated with inappropriate attendances at the emergency department of a tertiary hospital in Singapore. Singapore Med J 2020;61(2):75.
[9] Araz OM, Olson D, Ramirez-Nafarrate A. Predictive analytics for hospital admissions from the emergency department using triage information. Int J Prod Econ 2019;208:199–207.
[10] Whitt W, Zhang X. Forecasting arrivals and occupancy levels in an emergency department. Oper Res Health Care 2019;21:1–18.
[11] Harrou F, Dairi A, Kadri F, Sun Y. Forecasting emergency department overcrowding: A deep learning framework. Chaos Solitons Fractals 2020;139:110247.
[12] Kam HJ, Sung JO, Park RW. Prediction of daily patient numbers for a regional emergency medical center using time series analysis. Healthc Inform Res 2010;16(3):158–65.

[13] Ribeiro MHDM, da Silva RG, Mariani VC, dos Santos Coelho L. Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil. Chaos Solitons Fractals 2020;135:109853.
[14] Jusup M, Holme P, Kanazawa K, Takayasu M, Romić I, Wang Z, Geček S, Lipić T, Podobnik B, Wang L, et al. Social physics. Phys Rep 2022;948:1–148.
[15] Xu M, Wong T-C, Chin K-S. Modeling daily patient arrivals at emergency department and quantifying the relative importance of contributing variables using artificial neural network. Decis Support Syst 2013;54(3):1488–98.
[16] Singh S, Parmar KS, Makkhan SJS, Kaur J, Peshoria S, Kumar J. Study of ARIMA and least square support vector machine (LS-SVM) models for the prediction of SARS-CoV-2 confirmed cases in the most affected countries. Chaos Solitons Fractals 2020;139:110086.
[17] Cortez B, Carrera B, Kim Y-J, Jung J-Y. An architecture for emergency event prediction using LSTM recurrent neural networks. Expert Syst Appl 2018;97:315–24.
[18] Sudarshan VK, Brabrand M, Range TM, Wiil UK. Performance evaluation of emergency department patient arrivals forecasting models by including meteorological and calendar information: A comparative study. Comput Biol Med 2021;135:104541.
[19] 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.
[20] JuHyok U, Lu P, Kim C, Ryu U, Pak K. A new LSTM based reversal point prediction method using upward/downward reversal point feature sets. Chaos Solitons Fractals 2020;132:109559.
[21] Sagheer A, Kotb M. Time series forecasting of petroleum production using deep LSTM recurrent networks. Neurocomputing 2019;323:203–13.
[22] Law R, Li G, Fong DKC, Han X. Tourism demand forecasting: A deep learning approach. Ann Tour Res 2019;75:410–23.
[23] Wang P, Zheng X, Ai G, Liu D, Zhu B. Time series prediction for the epidemic trends of COVID-19 using the improved LSTM deep learning method: Case studies in Russia, Peru and Iran. Chaos Solitons Fractals 2020;140:110214.
[24] Chakraborty T, Ghosh I. Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis. Chaos Solitons Fractals 2020;135:109850.
[25] Altan A, Karasu S, Bekiros S. Digital currency forecasting with chaotic meta-heuristic bio-inspired signal processing techniques. Chaos Solitons Fractals 2019;126:325–36.
[26] Livieris IE, Pintelas E, Pintelas P. A CNN–LSTM model for gold price time-series forecasting. Neural Comput Appl 2020;32(23):17351–60.
[27] Perc M, Gorišek Miksić N, Slavinec M, Stožer A. Forecasting covid-19. Front Phys 2020;127.
[28] Al-Qaness MA, Ewees AA, Fan H, Abd El Aziz M. Optimization method for forecasting confirmed cases of COVID-19 in China. J Clin Med 2020;9(3):674.
[29] Zhang Y, Zhang J, Tao M, Shu J, Zhu D. Forecasting patient arrivals at emergency department using calendar and meteorological information. Appl Intell 2022;1–12.
[30] 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.
[31] Navares R, Aznarte JL. Deep learning architecture to predict daily hospital admissions. Neural Comput Appl 2020;32(20):16235–44.
[32] Kadri F, Abdennbi K. RNN-based deep-learning approach to forecasting hospital system demands: application to an emergency department. Int J Data Sci 2020;5(1):1–25.
[33] Parker CA, Liu N, Wu SX, Shen Y, Lam SSW, Ong MEH. Predicting hospital admission at the emergency department triage: A novel prediction model. Am J Emerg Med 2019;37(8):1498–504.
[34] ArunKumar K, Kalaga DV, Kumar CMS, Kawaji M, Brenza TM. Forecasting of COVID-19 using deep layer recurrent neural networks (RNNs) with gated recurrent units (GRUs) and long short-term memory (LSTM) cells. Chaos Solitons Fractals 2021;146:110861.
[35] Liu N, Xie F, Siddiqui FJ, Ho AFW, Chakraborty B, Nadarajan GD, Tan KBK, Ong MEH, et al. Leveraging large-scale electronic health records and interpretable machine learning for clinical decision making at the emergency department: Protocol for system development and validation. JMIR Res Prot 2022;11(3):e34201.
[36] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9(8):1735–80.
[37] Gers FA, Schmidhuber J, Cummins F. Learning to forget: Continual prediction with LSTM. Neural Comput 2000;12(10):2451–71.
[38] Sahoo BB, Jha R, Singh A, Kumar D. Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting. Acta Geophys 2019;67(5):1471–81.
[39] 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.
[40] 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.
[41] Sultana F, Sufian A, Dutta P. A review of object detection models based on convolutional neural network. Intell Comput: Image Process Based Appl 2020;1–16.
[42] Hoseinzade E, Haratizadeh S. CNNpred: CNN-based stock market prediction using a diverse set of variables. Expert Syst Appl 2019;129:273–85.
[43] Hiransha M, Gopalakrishnan EA, Menon VK, Soman K. NSE stock market prediction using deep-learning models. Procedia Comput Sci 2018;132:1351–62.
[44] Chimmula VKR, Zhang L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. Chaos Solitons Fractals 2020;135:109864.
[45] Kim J, Moon N. BiLSTM model based on multivariate time series data in multiple field for forecasting trading area. J Ambient Intell Humaniz Comput 2019;1–10.
[46] Lee SW, Kim HY. Stock market forecasting with super-high dimensional time-series data using ConvLSTM, trend sampling, and specialized data augmentation. Expert Syst Appl 2020;161:113704.
[47] Wang J, Li X, Li J, Sun Q, Wang H. NGCU: A new RNN model for time-series data prediction. Big Data Res 2022;27:100296.
[48] Liu X, Lin Z, Feng Z. Short-term offshore wind speed forecast by seasonal ARIMA-A comparison against GRU and LSTM. Energy 2021;227:120492.
[49] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521(7553):436–44.
[50] Telgarsky M. Benefits of depth in neural networks. In: Conference on learning theory. PMLR; 2016, p. 1517–39.
[51] Elfering A, Gerhardt C, Pereira D, Schenker A, Kottwitz MU. The monday effect revisited: a diary and sleep actigraphy study. Sleep Vigil 2020;4(2):167–76.
[52] DaGar S, Sahin S, Yilmaz Y, Durak U. Emergency department during long public holidays. Turk J Emerg Med 2014;14(4):165–71.
[53] Petsis S, Karamanou A, Kalampokis E, Tarabanis K. Forecasting and explaining emergency department visits in a public hospital. J Intell Inf Syst 2022;1–22.
[54] Chan SL, Ho AF, Ding H, Liu N, Earnest A, Koh MS, Chuah JS, Lau ZY, Tan KB, Zheng H, et al. Impact of air pollution and trans-boundary haze on nation-wide emergency department visits and hospital admissions in Singapore. Ann Acad Med Singapore 2020;49(2):78–87.
[55] Anderson H, Atkinson R, Bremner S, Marston L. Particulate air pollution and hospital admissions for cardiorespiratory diseases: are the elderly at greater risk? Eur Respir J 2003;21(40 suppl):39s–46s.