IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 11, 2020, accepted November 16, 2020, date of publication November 18, 2020,
date of current version December 3, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3039002

Deep Air Quality Forecasts: Suspended Particulate Matter Modeling With Convolutional Neural and Long Short-Term Memory Networks

EKTA SHARMA[1], (Member, IEEE), RAVINESH C. DEO[1], (Senior Member, IEEE), RAMENDRA PRASAD[2], ALFIO V. PARISI[3], AND NAWIN RAJ[1]

[1]School of Sciences, University of Southern Queensland, Springfield Central, QLD 4300, Australia
[2]Department of Science, School of Science and Technology, The University of Fiji, Lautoka, Fiji
[3]Centre for Applied Climate Sciences, School of Sciences, University of Southern Queensland, Toowoomba, QLD 4350, Australia

Corresponding authors: Ekta Sharma (ekta.sharma@usq.edu.au) and Ravinesh C. Deo (ravinesh.deo@usq.edu.au)

The work of Ekta Sharma was supported by the Research Training Scheme from the Australian Government and the School of Sciences, University of Southern Queensland.

ABSTRACT Public health risks arising from airborne pollutants, *e.g.*, Total Suspended Particulate (*TSP*) matter, can significantly elevate ongoing and future healthcare costs. The chaotic behaviour of air pollutants posing major difficulties in tracking their three-dimensional movements over diverse temporal domains is a significant challenge in designing practical air quality systems. This research paper builds a deep learning hybrid *CLSTM* model where convolutional neural network (*CNN*) is amalgamed with the long short-term memory (*LSTM*) network to forecast hourly *TSP*. The *CNN* model entails a data processer including feature extractors that draw upon statistically significant antecedent lagged predictor variables, whereas the *LSTM* model encapsulates a new feature mapping scheme to predict the next hourly *TSP* value. The hybrid *CLSTM* model is comprehensively benchmarked and is seen to outperform an ensemble of five machine learning models. The efficacy of the *CLSTM* model is elucidated in model testing phase at study sites in Queensland, Australia. Using performance metrics, visual analysis of *TSP* simulations relative to observations, and detailed error analysis, this study ascertains the *CLSTM* model's practical utility for air pollutant forecasting systems in health risk mitigation. This study captures a feasible opportunity to emulate air quality at relatively high temporal resolutions in global regions where air pollution is a considerable threat to public health.

INDEX TERMS Air quality forecasting, convolutional neural networks, deep learning, long short-term memory networks.

NOMENCLATURE

APF	Air Pollutant Forecasting.
AQ	Air Quality.
DL	Deep Learning.
$\mu g/m^3$	Micrograms per cubic metre.
$PM_{2.5}$	Fine particles with size 2.5 mm or less.
TSP	Total Suspended Particulate Matter up to about 100 μm In Diameter.
TSP_i^{FOR}	Forecasted *TSP* for i^{th} observation.
RELU	Rectified Linear Units.
WI	Willmott Index of agreement (*WI*)
ADAM	Adaptive Moment Estimation.
MAPE	Mean Absolute Percentage Error (%).
E_{NS}	Nash–Sutcliffe Efficiency.
RELU	Rectified Linear Units.
PM_{10}	Coarse particles between 2.5 And 10 mm.
CLSTM	Deep Learning hybrid Convolutional Long Short-term Memory Neural Network.
TSP_i^{OBS}	Observed *TSP* for i^{th} observation.
L	Legates and Mccabe Index.
r	Pearson's correlation coefficient.

The associate editor coordinating the review of this manuscript and approving it for publication was Asad Waqar Malik.

I. INTRODUCTION

As urbanisation progresses, air pollution is becoming an alarming environmental and societal concern. Besides regular monitoring efforts, there is a rising demand for short-term air pollutant forecasting (*APF*) system. This system can benefit

governments in its health policymaking, traffic control in times of heavy pollution, and protecting vulnerable factions of the society (*e.g.* senior citizens, people with health ailments, pregnant ladies, and children).

One critical environmental pollutant that has an indiscriminate emission footprint is particulate matter (*PM*). The diversity and complexity of *PM* movements make the ongoing analysis and forecasting of this health hazard a critically challenging task. This atmospheric property has been extensively studied (*e.g.* [1], [2]). *PM* constitutes $PM_{2.5}$ (Particulate Matter *2.5*, or fine pollutants $< 2.5\ \mu m$ (micrometres)), PM_{10} (Particulate Matter *10,* or coarse pollutants $(2.5 - 10)\ \mu m$), and total suspended particulate matter (*TSP or SPM*) *i.e.* all airborne particles up to $100\ \mu m$ in diameter [3]. *TSP*, measured in $\mu g/m^3$, is the subject of this paper, built on our earlier study [4] where we modelled visibility reducing particles and *PM* using conventional artificial intelligence models, albeit without considering *TSP*. This research, therefore, aims to design an *APF* system that predicts *TSP* responsible for recurrent healthcare costs and increased nuisance through the soiling of property and materials. The primary source of *TSP*, being combustion (*e.g.*, engines, bushfires, dust, mining, and industrial processes), is currently rising in many nations. There are very few studies globally that studied the acute effects of *TSP* and mortality [3], [5]. Therefore, new research is crucial for modelling the behavior of this health hazard.

In recent years, the forecasting of air pollutants has been accomplished through two methods: first, dynamic or physical models [6], and second, data-based statistical or artificial intelligence models [7] are considered. Physical models provide good accuracy of the physical processes, however, some studies report a significant model-error and a need for long run-times that make the model difficult to implement over a short-term horizon [8]. Fortunately, neural network models can address such issues as they only require proper datasets to speed up the learning and model convergence. Owing to the limitations of physics-based models, an *APF* system based on artificial intelligence can be a viable option to produce better results [9]. In atmospheric predictions, deep learning (*DL*) has attracted considerable attention. Many studies [10] are showing its ability to attain high forecasting precision than its earlier counterpart, or non-*DL* models. Numerous works have implemented *DL* in a diverse range of applications such as solar radiation [11], [12], pain intensity estimation [13], and seizure diagnosis [14]. In these studies, and the others, *DL* was commended for its superior capability to handle complex data (e.g., *TSP*) and approximation through stochastic variables analysis with a nonlinear feature mapping capability.

To develop an *APF* system, this research adopts a convolutional neural network (*CNN*). This is a renowned *DL* algorithm employing efficient multistage architecture through convolution, pooling, and fully connected layers for effective task-dependent and non-handcrafted data attribute representation [15]. Further improvements can be achieved through a secondary *DL* architecture based on long short-term memory (*LSTM*) network [16]. *LSTM's* key merits are that it can

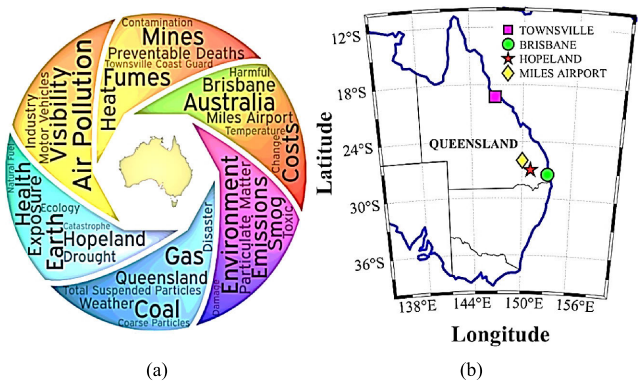

FIGURE 1. (a) A word crunch showing the problems caused by air pollution in Australia. (b) Study sites in Queensland Australia where the proposed *CLSTM* was implemented.

resolve to vanish gradient issues and explore sequential data relationships through unique (input, forget, output) gates. The amalgamation of *CNN* with *LSTM* is fast becoming a popular area of research in *AQ* and there are some important global studies such as [17]–[19]. However, the construction of the hourly *APF* system for *TSP* and especially for Australia is yet to be explored. Following the aforesaid, there appear to be gaps in the literature. *Firstly,* there is rather fragmentarily available literature involving *DL* algorithms for *TSP* prediction, and *Secondly*, there appear to be gaps in *TSP* prediction application in Australia despite a handful of studies performed elsewhere, *e.g. TSP* concentration models in China sea with back-propagation neural network (*BPNN*) [20], multi-layer filtration system [21], modelling *TSP* with light scattering [22], and with coulter sensors [23]. However, none of these studies have forecasted *TSP w.r.t* air quality. To address such issues, this research aims to build an *APF* system for short-term (hourly) *TSP* forecasts using the *LSTM* algorithm as a versatile model integrated with *CNN* as a feature extraction framework, as with $PM_{2.5}$ [24] and other studies [9], [25]. Despite air pollution causing 3000 premature deaths in Australia with a staggering annual health expense of *AUD* 24.3 billion, there is a dearth of research [26]. Recent spine-tingling Black Summer bushfires across Southeastern Australia (July 2019–February 2020) [27], [28] and ancient dust-storms registered severe air pollution episodes. Australia is an arid continent where the rising global temperature and bushfires, health, and environmental challenges illustrated graphically (Fig. 1 a) are considered serious [29]. An optimal *AQ* target is a critical issue but this is hindered by ineffective policies that warrant the development of a real-time and robust *APF* methodology [4]. Considering this, the novelty of this paper is as follows:

i) A practical study is presented to generate a computationally efficient architecture with the *DL* hybrid *'CLSTM'* model. Our approach aims to address gaps in hourly *TSP* modelling for a user-friendly *APF* system.
ii) *Firstly,* our model design phase employs an ensemble of competing machine learning models: Random Forest (*RF*), Volterra, M5 model tree, and Multiple

Linear Regression (*MLR*) to forecast hourly *TSP*. *Secondly,* the *LSTM* algorithm is applied to an ensemble of forecasted *TSP* to compare its results with the final output. Thereafter, a three-layered *CNN* model robustly extracts remaining data features *i.e.*, statistically significant antecedent inputs from 'four' (excluding *LSTM*) benchmarked models. The last layer analyses all features involving independent *LSTM* fusion to finally forecast the next hour *TSP* (μg/m^3). This multi-modeling overcomes the inherent deficiencies in single models [30].

iii) The *CLSTM* necessitates the data fusion through a refining process by *CNN* and features mapping by *LSTM*. Extensive evaluation through careful selection of last hour's meteorological variables over hotspots in Queensland, Australia, shows reliable forecasts.

iv) *CLSTM's* efficacy is explored by statistical metrics, visual analysis of forecasted and observed *TSP*, and comprehensively benchmark against an ensemble of five machine learning models including *LSTM*.

v) The hourly predictions of *TSP*, which is near real-time, can help combat public health issues through proactive advisory, planning, and implementing a regulatory *AQ* system, and making this research significantly unique for global applications for human health benefits.

II. THEORETICAL OVERVIEWS

We now provide a brief overview of the objective model (*i.e.*, *CNN* & *LSTM*). The theoretical explanation of standalone models *i.e. MLR* [31], M5 model tree [32], Volterra [30], and Random Forest [33] are presented elsewhere as these are well-known methodologies.

A. FEATURE EXTRACTION: CONVOLUTIONAL NEURAL NETWORK

In this research, *CNN* constructs the proposed *APF* system, denoted as *CLSTM*, which is applied for hourly *TSP* prediction. *CNN*'s are relatively successful Feed Forward Neural Networks [34]. However, few studies *e.g.* [35] have applied *CNN* for *AQ* research. A typical C*NN* architecture consists of Convolutional layers (*CON*) discovering the data patterns. This transforms local relationships in input features or images using a kernel. A Pooling layer (*POOL*) reduces the target variable dimension while a Fully connected layer (*FC*) generates the probability of each category implied in initial input data [36]. If a = activation function, W^f = kernel's weight with feature map f^{th}, and $*$ = an operator of convolutional process, each convolutional layer extracts *TSP* pattern with a lagged matrix, expressed mathematically as:

$$h_i j^k = a((W^f * x)_i j + b_k) \tag{1}$$

It should be noted that *CNN*'s are computationally intensive in extracting hidden nonlinear features. These features can help a *CNN* model to create filters representing the data patterns [37]. This paper aims to simplify the modelling technique, mainly to satisfy real-time usage for hourly *TSP* architecture. A one-dimensional (1-D) *CON* operator directly forecasts 1-D *TSP* data with a grid search being used to select the channels through three *CON* layers. *Adam* is an optimisation algorithm and *ReLU* is an optimisation algorithm such that *ReLU* is:

$$f\,(x) = \max\,(0, x) \tag{2}$$

B. LONG SHORT-TERM MEMORY NETWORK: TIME SERIES PREDICTION

LSTM is a Recurrent Neural Network (*RNN*) [38] with the ability to learn long-range dependence between target and input variables. Hence *LSTM's* are becoming useful in forecasting time series variables, especially in *AQ* research as the objective is to consider dependence on a forecast horizon suited for sequential prediction whilst also evading the gradient decay issues. The memory block of *LSTM* has input, output, and forget gates, assisting the update of information flow, thus making it an excellent choice for *TSP* modelling. This can continuously update the next forecasted value [39]. Successful usage of *LSTMs* are language modelling [40], speech recognition [41], and *AQ* research [42]. The calculations are as follows [9]:

I. If S_{n-1} = Last hidden state, I_n = new input, F_n = forget gate, W_m = Weight matrices, b_m = bias vector, i_n = input gate, $\sigma(\ldots)$, $\tanh(\ldots)$ = Activation functions (Logistic, sigmoid, hyperbolic), the cell state is represented by:

$$F_n = \sigma\,(W_m.\,(S_{n-1}, I_n) + b_m) \tag{3}$$

A candidate cell state δ_n decides what information will be stored, scaled by i_n –

$$\tanh\,(W_c.\,(S_{n-1}, I_n) + b_c) \tag{4}$$

$$i_n = \sigma\,(W_i.\,(S_{n-1}, I_n) + b_i) \tag{5}$$

II. Cell state δ_n combines the earlier state and the present state (δ_{n-1}, δ_n). Here δ_{n-1} is scaled by i_n and δ_{n-1} by F_n:

$$\delta_n = (F_n * \delta_{n-1} + I_n * \delta_n) \tag{6}$$

III. Finally, for the output process, (''output gate'' θ_n decide the output state $\mathfrak{C}_n$. Here θ_n is filter for output S_n :

$$\theta_n = \sigma(W_\theta.(S_{n-1}), I_n) + b_\theta) \tag{7}$$

$$S_n = \theta_n.tanh(\delta_n) \tag{8}$$

III. MATERIALS AND METHOD

A. RESEARCH AREA

To appraise CLSTM, we utilize hourly air pollutants or TSP (μg/m^3) for Queensland (Qld). Table 1 (a-b) describes the study sites and all relevant data. *Qld* is the second largest state in Australia, exhibiting a soaring rate of greenhouse gas emissions per capita [43]. It has nine (out of ten) worst mines that generate *PM*, causing major respiratory issues with cancer cases [44]. A spatial picture of the study sites is illustrated in Fig. 1 (b). The *Qld*-based

TABLE 1. (a) Geographic description (b) Data segregation of study sites.

Station Name	*Location*		
	Longitude (°E)	*Latitude (°S)*	*Elevation (m)*
Brisbane	153.08	27.46	9.0
Townsville	146.81	19.25	9.0
Hopeland	151.04	27.04	316.0
Miles Airport	150.16	26.81	302.0

(a)

(b)

Station Name	*Data points*	*Training*			*Validation*			*Testing*		
		Period	*Points*	*%*	*Period*	*Points*	*%*	*Period*	*Points*	*%*
Brisbane	35064	01-Jan-2015 to 31-Dec-2017	26304	75.01	01-Jan-2018 to 30-June-2018	4344	12.38	01-July-2018 to 31-Dec-2018	4416	12.59
Townsville	35064	01-Jan-2015 to 31-Dec-2017	26304	75.01	01-Jan-2018 to 30-June-2018	4344	12.38	01-July-2018 to 31-Dec-2018	4416	12.59
Hopeland	34752	14-Jan-2015 to 31-Dec-2017	25992	74.79	01-Jan-2018 to 30-June-2018	4344	12.50	01-July-2018 to 31-Dec-2018	4416	12.70
Miles Airport	30678	02-Jul-2015 to 31-Dec-2017	21918	71.45	01-Jan-2018 to 30-June-2018	4344	14.16	01-July-2018 to 1-Dec-2018	4416	14.39

TABLE 2. Descriptive statistics of *TSP* ($\mu g/m^3$) for each study site.

Objective Variable	*Study Site*	*Maximum*	*Minimum*	*Mean*	*Median*	*Variance*	*Skewness*	*Kurtosis*	*Standard-Deviation*
TSP ($\mu g/m^3$)	Brisbane	2568.4	0.10	24.83	21.4	556.48	41.12	3941.85	23.59
	Townsville	439.9	0.10	25.92	23.2	221.71	3.19	36.46	14.89
	Hopeland	992.3	0.01	17.87	11.6	990.36	12.75	246.42	31.47
	Miles Airport	3709.1	0.40	36.56	20.5	6310.71	13.23	325.94	79.44

air monitoring hotspots are Brisbane, Townsville, Hopeland, and Miles Airport. The Brisbane station monitors *PM* (and *TSP*) which is potentially relevant to emissions from rail wagons transferring coal to the Port of Brisbane. Townsville station began in 2007 in the form of a dust monitoring program by considering community concerns regarding dust impacts from the Port of Townsville. Rural sites: Hopeland and Miles Airport are responsible for assessing *AQ* near an area of intensive coal seam gas productions. Considering the need to develop forecast models for *TSP* especially for these important hotspots, the study designed the proposed *CLSTM* model. Table 2 displays air pollutant *TSP's* inferential statistics. Data were acquired from the *Qld* Department of Environment and Science. Data are continuously evaluated for quality through preliminary analysis. Table 1 (b) shows missing data ('×') caused by equipment servicing and other instrumental failures. Following statistical procedures, the hourly mean calendar values were used to replace missing data [45].

B. DEEP HYBRID CLSTM MODEL ARCHITECTURE

Hourly TSP data were used to develop six forecast models shown in the overall schematic (Figure 2) of the proposed deep learning (DL) hybrid CLSTM model. The objective model is compared to the DL model LSTM, including non-DL versions: Volterra, Random Forest, MLR, and M5 model tree. All models were developed on Windows 10 platform Intel®i7 Generation 9 @ 3.7 gigahertz processing unit, 16 GB memory, Python programming language, and freely available open-source libraries (i.e., Keras [46], Tensor Flow [47], and Scikit-learn [48]). To develop a hybrid CLSTM model, data were firstly partitioned; and although there is no consensus over data partitioning ratios [49]; this issue is a critical consideration as it affects the feasibility and capability of the model. In absence of any designated rule [50], the TSP data were initially divided into training sets (01-January-2015 to 31-December-2017) for Brisbane and Townsville (i.e. 75%). For Hopeland, and Miles Airport the period considered was 14-January-2015 to 31-December-2017 (i.e. 75%), and 02-July-2015 to 31-December-2017 (i.e. 72% or 30,678) when these stations started their AQ monitoring. This data division ensured a universally representative hourly-step horizon. Table 1(b) shows that a six-monthly period for validation (01-January-2018 to 30-June-2018) and testing (01-July-2018 to 31-December-2018) purposes for all stations. This is consistent with published literature that emphasises the criticality of partitioning into subsets before model construction to avoid leakage of validation and training data over the upcoming testing subset, thereby introducing a testing bias [51]. The sample data from the training set provided model estimation through hyper-parameters hence the validation set became an essential component of modelling. Table 2 enumerates the descriptive statistics of measured AQ, and Table 3 discusses the model input variables used for hourly TSP ($\mu g/m^3$) forecasting. Following earlier literature,

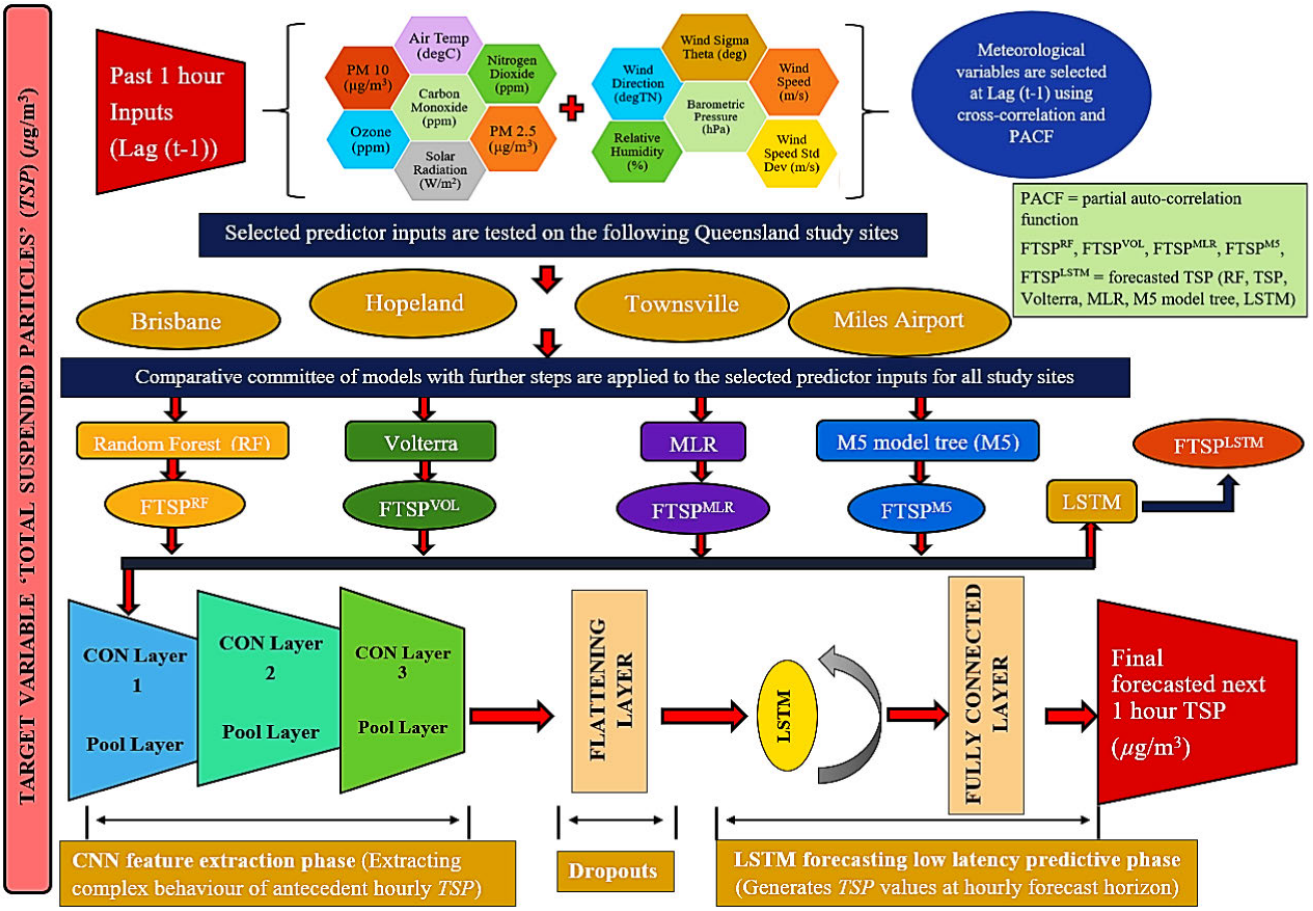

FIGURE 2. Schematic of *APF* system using deep learning hybrid model (*CLSTM*).

the chaotic nature of AQ requires a normalisation process before modelling to conform all data to be within [0, 1] [52]:

$$TSP_{NORM} = (TSP - TSP_{MIN})/(TSP_{MAX} - TSP_{MIN}) \quad (9)$$

In (9), TSP_{MIN}, and TSP_{MAX} = the minimum and maximum values of *TSP*, respectively, while TSP_{NORM} = the normalised *TSP*. In Table 3, Wind (*W*), Direction (*D*), Speed (S), Air Temperature (*AT*), Relative Humidity (*RH*), Barometric pressure (*BP*), (NO_2), Carbon monoxide (*CO*), and Solar Radiation (*RAD*). The next step in model design is illustrated in Fig. 3 (a)-(d). It identifies cross-correlation coefficients (r_{cross}) investigating the co-variance between hourly *AQ* data *vs.* respective predictor variables used to build a hybrid *CLSTM* for all study sites. The blue line indicates a 95% significance boundary. Fig. 4 (a)–(d) utilises partial autocorrelation function (*PACF*) to deduce the correlation of *TSP* series with its own historical lagged values, regressing the shifted series to determine these correlations. Successive time-shifted predictors created by this method denoted as *TSP* $(t - n)$ where n is the value of the lag (*e.g.*, $n = 1$ in Fig 4) reveals good correlation with past *TSP* at all sites. Through this method, the significant lagged *TSP* series is then employed to build the proposed *CLSTM* model. In Table 4, the optimal framework of *CLSTM* (boldfaced in **red**) *vs.* the comparative models is illustrated. Notably, a grid search was adopted to select the hyperparameters and corresponding

TABLE 3. Model input variables. × = data is not monitored.

Variable	*Units*	*Brisbane*	*Townsville*	*Hopeland*	*Miles*
WD	(°*TN*)	✓	✓	✓	✓
WS	(*m/s*)	✓	✓	✓	✓
WST	(°)	✓	✓	×	×
WSSD	(*m/s*)	✓	✓	×	×
AT	(°*C*)	✓	✓	✓	✓
RH	(*%*)	✓	×	✓	✓
BP	(*hPa*)	✓	×	×	×
PM_{10}	(μg/m^3)	×	✓	×	✓
$PM_{2.5}$	(μg/m^3)	×	×	×	✓
O_3	(*ppm*)	×	×	✓	✓
NO_2	(*ppm*)	×	×	✓	✓
CO	(*ppm*)	✓	✓	✓	✓
RAD	(W/m^2)	✓	✓	✓	✓

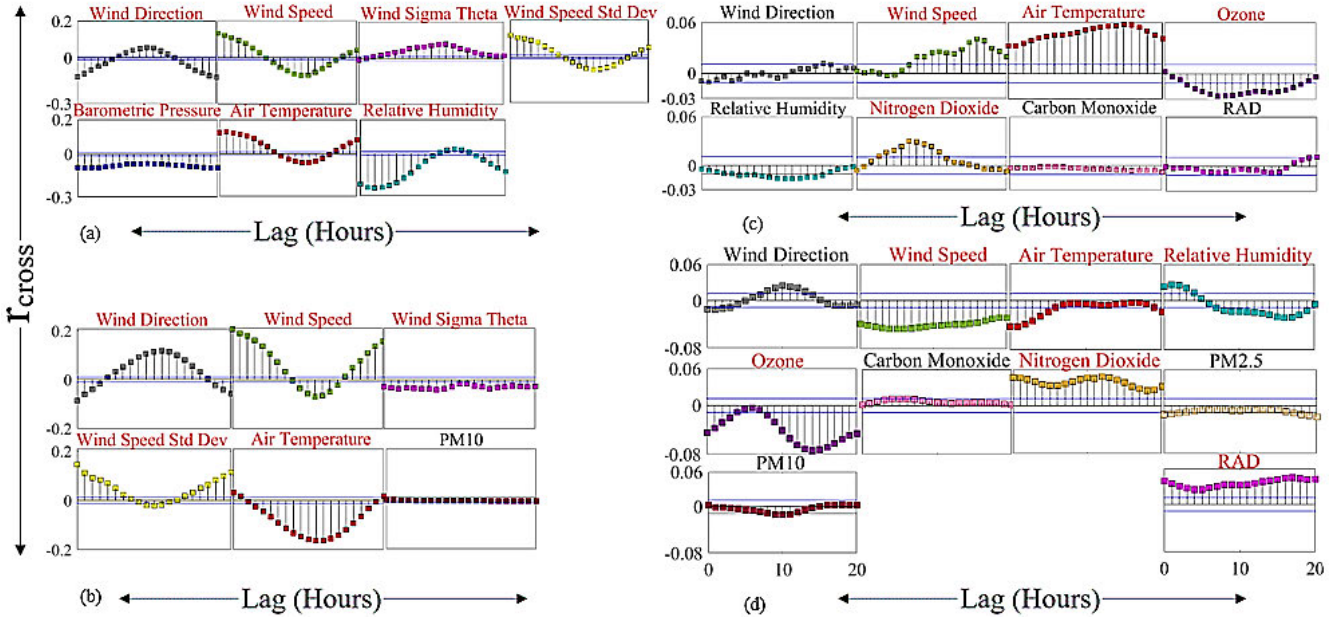

FIGURE 3. Cross-correlations (r_{cross}) investigating co-variance between the *AQ vs.* predictor variables used to build the proposed *APF* system using a hybrid *CLSTM* model. Note: The selected variables are captioned in red (a) Brisbane (b) Townsville (c) Hopeland, (d) Miles Airport.

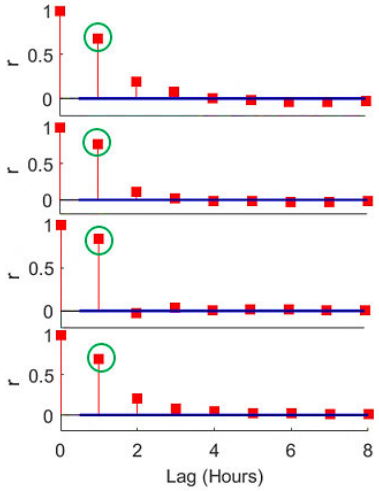

FIGURE 4. Partial autocorrelation of TSP time-series. The green symbol shows the most significant lagged TSP used to develop the proposed APF system using the hybrid CLSTM model. (a) Brisbane (b) Townsville (c) Hopeland, (d) Miles Airport.

optimal value range to finally reach an optimum framework for best feature extraction. The starting numbers are determined as per published studies conducted for *DL* [9], [12]. The model's improvement was monitored by the successive addition of numbers until we reach an optimal set. For instance, as shown in Table 4, when *CLSTM* was tested for a good batch size (from 400 as it is big data) and it continued till the performance became optimal (*i.e.* 2000 epochs and 500 batch size). It is noteworthy that the optimal number of hidden neurons plays a significant role in determining the most feasible model architecture. In this study, it was ensured that the grid search enabled the model to correctly learn the predictor data patterns, and thus, avoided issues like overfitting that can occur in cases when a significant model framework does not converge in a given run time or a small framework might lack the appropriate degrees of freedom to fully extract data patterns [53]. The lowest training *MAE* or *RMSE*, which was obtained by a sequence of hidden neurons in gradual steps was adopted to determine the optimal model architecture [54]. After trialling SoftMax, tangent hyperbolic, and sigmoid functions, '*ReLU*' was found to be the optimal activation function attaining the best *CLSTM* performance. Hyper-parameters of all tested models were chosen through a grid search procedure, which could impose a huge time factor and significant computing costs associated with the machine learning process. Each model search generally takes 10∼11 hours with the computation training and testing time to get reduced to $<$ 20 min after deducing the optimal parameters [9]. Notwithstanding this, the training data size is known to affect the cost and hyperparameter selection [55]. The deep learning hybrid model *CLSTM* utilises a pooling layer to overcome overfitting issues in its training phase, helping minimise and control the parameters and computations involved. It should be clarified that the inputs of *CLSTM* are the hourly *TSP*'s lagged matrix; however, the output is the forecasted *TSP*. The next *i.e.* the fourth layer in the *CLSTM* framework is the *LSTM* architecture itself positioned to analyse the features and later pass those in forecasting *TSP* (μg/m^3) values for the next hour. As enunciated in grid search after hyper-parameter selection, the study utilises the following properties [9]:

- **Least Square Error and Absolute Deviation**(L1, L2-regularisation): The sum of absolute differences and square of differences between observed and forecasted

TABLE 4. Optimal architecture (in red for CLSTM) for *TSP*(μg/m^3).

Model	*Hyper-parameters*	*Grid Search for Optimal set*
CLSTM	Epochs	[300, 500, 1000, **2000**]
	Activation function	[SoftMax, tanh, **ReLU**, sig]
	Batch size	[400, **500,** 800, 750, 1000]
	Pooling size, padding	[**2**], [same]
	Convolution Layer 1	[250, 200, 150, 100, **80**]
	Convolution Layer 2	[**20**, 40, 50, 60, 70, 80]
	Convolution Layer 3	[30, 10, 15, 20, **5**]
	LSTM layer (L1)	[30, 40, **50**, 60, 100, 150]
	Optimiser, Dropout	[**Adam**], [**0.1**]
LSTM	Epochs	[50, 100, 500, 1000, 2000]
	Activation function	[SoftMax, tanh, ReLU, sig]
	Batch Size	[300, 500, 800, 750, 1000]
	Optimiser, Drop rate	[Adam], [0.1, 0.2]
	LSTM filter	[50, 60, 100, 200]
Random Forest	Number of Trees	[50, 100, 200, 400, 500, 800]
	Leaf, Foot, Surrogate	5, 1, 'on'
Volterra	Threshold, Regressor	[0,1]
	A, n-start	[m X n] matrix [m>n], [>=1]
MLR	Intercept	[0,1]
	c	[0.0238, 0.758,0.183,0.157]
	α_1	[1.207,0.005,0.008, -0.021]
	α_2	[0.820,0.102, -0.087,0.267]
	α_3	[0.004, -0.012, -0.655, -0.345]
M5 model tree	Threshold, Smoothing	0.05, 15
	Minimum cases, rules	5, 20

The Architecture of the Backpropagation Algorithm	
Beta, β_1, β_2	0.990
Alpha, α	0.001
Epsilon, ε	0.0000001
β_1, β_2 = 1st, 2nd moment estimates exponential decay rate	
α = Learning rate, ε = Small number to prevent zero division	

TSP is minimised through penalisation parameters. Generally, L1, L2 parameters penalises the huge parameter values, thus reducing the models' nonlinearity.

- **Dropout:** is a regularisation adopted to improve the training performance and reduce overfitting. For all iterations, dropout picks up a neuron fraction between [0,1] (hyper-parameter). In this study, dropout = 0.1.
- **Activation Function:** SoftMax, tanh, sigmoidal were tested with Rectified Linear Unit (*ReLU*) as most optimal.
- **Early Stopping:** Kera's *DL* library eliminates overfitting [56] by setting the mode to "minimum" and patience to "45". Here, *ES* typically halts training when validation loss stops, decreasing the number of epochs, specified by the patience term.

Summarising these, the hybrid *CLSTM* was obtained in three distinct stages: **Selection:** Predictors based on meteorological variables at a lag of (*t–1* or using features of the past hour) were selected using r_{cross} and *PACF*. **Validation & Testing:** Validation of the selected inputs and testing on *Qld* sites was carried out through these models: Volterra, Random Forest, M5 model tree, and *MLR*. *LSTM* was applied to the forecasted *TSP* (μg/m^3) result. **Feature extraction & low latency predictive phase:** The three-layer *CNN* feature extraction phase was employed on forecasted *TSP* (four models). The fourth layer was the low latency predictive fusion phase with an independent *LSTM* on the flattened layer forming a single connected layer, giving the final forecasted next hour *TSP*, with results evaluated *via* equations (10) – (15).

C. MODEL PERFORMANCE CRITERIA

This subsection provides a rigorous evaluation of the hybrid CLSTM relative to its counterpart comparative models adopting a wide range of statistical criteria based on Pearson's correlation (r), Root-Mean-Square-Error (RMSE; μg/m^3), Mean Absolute Percentage Error (MAPE; %), Mean Absolute Error (MAE; μg/m^3), Willmott's Index of agreement (WI), Nash–Sutcliffe Efficiency (E_{NS}), and Legates and McCabe Index (L) [50], [57], [58] whose mathematical equations are:

I) Mean absolute error

$$(MAE) = \frac{1}{N}\sum_{i=1}^{N}\left|\left(\boldsymbol{TSP^{FOR} - TSP^{OBS}}\right)\right| \quad (10)$$

II) Root mean square error

$$(RMAE) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(TSP_i^{FOR} - TSP_i^{OBS}\right)^2} \quad (11)$$

III) Pearson's correlation coefficient (r) (12), as shown at the bottom of the next page

IV) Willmott Index of agreement (*WI*)

$$1 - \left[\frac{\sum_{i=1}^{N}\left(TSP_i^{FOR} - TSP_i^{OBS}\right)^2}{\sum_{i=1}^{N}\left(\left|TSP_i^{FOR} - \overline{TSP_i^{OBS}}\right| + \left|TSP_i^{OBS} - \overline{TSP_i^{OBS}}\right|\right)^2}\right] \quad (0 \le WI \le 1) \quad (13)$$

V) Nash–Sutcliffe Efficiency (E_{NS})

$$1 - \left[\frac{\sum_{i=1}^{N}\left(TSP_i^{FOR} - TSP_i^{OBS}\right)^2}{\sum_{i=1}^{N}\left(TSP_i^{OBS} - \overline{TSP_i^{OBS}}\right)^2}\right] \quad (\infty \le E_{NS} \le 1) \quad (14)$$

VI) Legates and McCabe Index (L)

$$1 - \left[\frac{\sum_{i=1}^{N}\left|TSP_i^{OBS} - TSP_i^{FOR}\right|}{\sum_{i=1}^{N}\left|TSP_i^{OBS} - \overline{TSP_i^{OBS}}\right|}\right] \quad (\infty \le L \le 1) \quad (15)$$

where TSP_i^{FOR}, TSP_i^{OBS} = forecasted and observed TSP for i^{th} observation, N = Total number, $\overline{TSP_i^{FOR}}$, $\overline{TSP_i^{OBS}}$ = mean forecasted and observed *TSP*. The study considered that a Gaussian error distribution is likely to imply *RMSE* to be a more appropriate measure of model accuracy compared to the *MAE* [59]. Note that a better performance is attained for *WI* and E_{NS} close to unity [50], [57], [58]. However, the value of *L* (a refined WI) is used to penalise model error more strictly, and so, it is a considerably robust metric, especially for large and relatively complex datasets (*e.g.*, *TSP*).

TABLE 5. Testing performance of *CLSTM vs.* Other competing models. (a) Brisbane, (b) Townsville, (c) Hopeland, (d) Miles airport.

	Models	r	MAE $\mu g/m^3$	$RMSE$ $\mu g/m^3$	$MAPE$ %	WI	E_{NS}	L
(a)	**CLSTM**	**0.93**	**3.54**	**13.97**	**8.43**	**0.96**	**0.86**	**0.80**
	LSTM	0.82	4.69	23.31	10.20	0.83	0.63	0.73
	RF	0.90	15.51	40.74	22.01	0.93	0.81	0.64
	Volterra	0.86	7.91	19.24	25.12	0.95	0.74	0.54
	MLR	0.86	8.13	20.27	27.08	0.89	0.71	0.53
	M5	0.38	11.66	46.71	36.07	0.55	0.52	0.33
(b)	**CLSTM**	**0.83**	5.40	**7.44**	**17.37**	0.91	**0.69**	**0.51**
	LSTM	0.81	5.12	7.94	19.99	0.89	0.64	0.45
	RF	0.79	5.65	8.73	26.74	0.77	0.61	0.44
	Volterra	0.76	**4.63**	9.05	21.56	**0.97**	0.54	0.42
	MLR	0.76	5.34	8.63	27.51	0.84	0.54	0.43
	M5	0.46	9.02	15.69	43.93	0.65	0.37	0.36
(c)	**CLSTM**	**0.67**	**7.98**	**21.97**	49.84	**0.79**	**0.43**	**0.35**
	LSTM	0.55	8.04	24.23	**38.44**	0.66	0.29	0.34
	RF	0.45	8.74	26.11	57.87	0.33	0.18	0.29
	Volterra	0.44	9.96	27.45	43.26	0.70	0.10	0.19
	MLR	0.44	8.91	25.97	59.78	0.51	0.19	0.27
	M5	0.37	9.80	29.62	45.80	0.54	0.10	0.20
(d)	**CLSTM**	**0.85**	**6.60**	**23.11**	**32.52**	**0.87**	**0.69**	**0.61**
	LSTM	0.77	7.76	26.66	53.16	0.81	0.58	0.53
	RF	0.68	9.04	30.27	37.18	0.71	0.45	0.46
	Volterra	0.65	9.31	31.94	35.03	0.82	0.39	0.44
	MLR	0.45	12.01	49.35	113.57	0.62	0.44	0.28
	M5	0.64	13.81	32.03	55.55	0.76	0.39	0.17

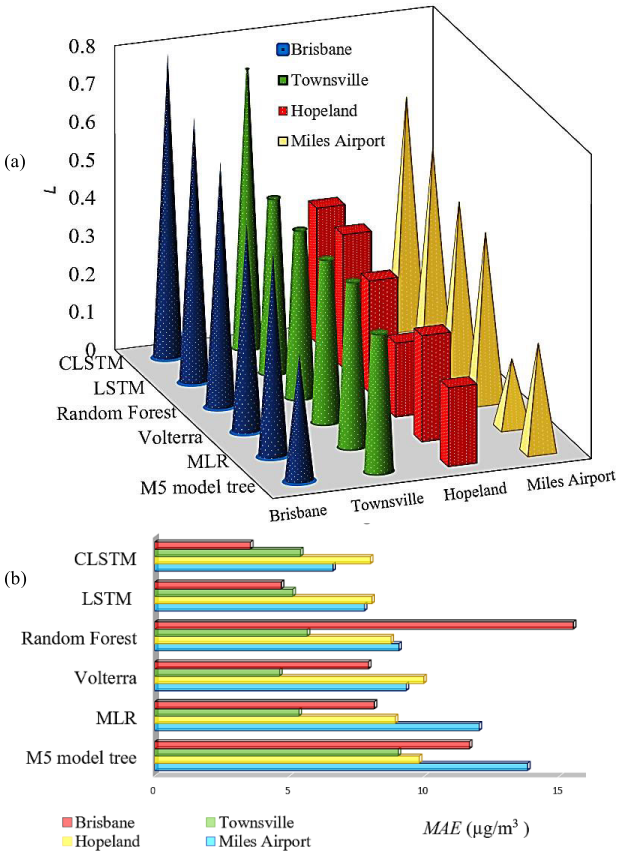

FIGURE 5. (a) Testing performance of *CLSTM vs.* the five other competing models evaluated using Legates & McCabe Index (*L*). (b) 3D-Bar graph of the mean absolute error *(MAE).*

IV. RESULTS AND DISCUSSIONS

This section provides empirical results to assess model performance, demonstrating the effectiveness of the newly designed *APF* system using a hybrid *CLSTM* approach. The results of hybrid *CLSTM* are benchmarked with the five other competing approaches; *e.g.*, *LSTM*, *RF*, Volterra, M5 model tree, and *MLR*. Further model appraisal has been achieved through several model evaluation metrics as described by (10)–(15). Table 5 (a)–(d) shows the predictive performance for all models evaluated in the testing phase. Comparing the results of the study site Brisbane plotted in Fig. (5–9) for TSP, this work attained the most precise forecasts for the case of the hybrid *CLSTM w.r.t* all six statistical metrics *(highest* $r \approx 0.93$, *lowest* $RMSE \approx 13.97 \mu g/m^3$, $MAE \approx 3.54 \mu g/m^3$, *highest* $WI \approx 0.96$, $L \approx 0.80$, and $E_{NS} \approx 0.86$*)* in comparison with statistical metrics of the other models. For instance, $r \approx [0.38\text{-}0.82]$, $MAE \approx [4.69\text{-}15.51] \mu g/m^3$, $MAPE \approx [10.20\text{-}36.07]\%$, $RMSE \approx [19.24\text{-}46.71]\ \mu g/m^3$, $WI \approx [0.55\text{-}0.95]$, $E_{NS} \approx [0.52\text{-}0.81]$, and $L \approx [0.33\text{-}0.73]$ for the comparative ensemble of models, where the value before [-] is the lower bound and the value after [-] is the upper bound of a metrics. For a complete understanding of *CLSTM,* we evaluate the test performance with all benchmark models (*i.e.*, *LSTM*, *RF*, M5 Model Tree, and *MLR*) through the Legates and McCabe Index (L) and a 3D-bar graph of the mean absolute error (*MAE*) of hourly *TSP* forecasts, *i.e.*, Fig. 5(a), (b). The results show the highest 'L' and lowest errors, highlighted by the performance metrics. That is, we get the following results: Brisbane ($MAE \approx 3.54\ \mu g/m^3$), Townsville ($MAE \approx 5.40\ \mu g/m^3$), Hopeland ($MAE \approx 7.98\ \mu g/m^3$), and Miles Airport ($MAE \approx 6.60\ \mu g/m^3$), all of which accord with excellent performance of the model in terms of current literature [60]. Our deep learning hybrid model: *CLSTM* shows the highest L ($\approx 0.80, 0.51, 0.35, 0.61$) for all study sites when compared with the rest of the model ensembles. It should be noted that this is the most stringent performance measure and thus, is an indicator of the superior performance of the proposed *CLSTM* model [61]. Furthermore, the 'p' value of the *CLSTM* model has been tested for all stations at significance level 0.05, 0.01, and 0.1. The p-value is < 0.00001 for all study sites, which indicates the result is significant at $p < 05$, 0.1, and 0.01. Testing performance of *CLSTM* model An alternative account of

$$\frac{\sum_{i=1}^{N}\left(TSP_i^{OBS}-\overline{TSP^{OBS}}\right)\left(TSP_i^{FOR}-\overline{TSP^{FOR}}\right)}{\sqrt{\sum_{i=1}^{N}\left(TSP_i^{OBS}-\overline{TSP^{OBS}}\right)^2}\sqrt{\sum_{i=1}^{N}\left(TSP_i^{FOR}-\overline{TSP^{OBS}}\right)^2}} \quad (-1 \le r \le 1) \tag{12}$$

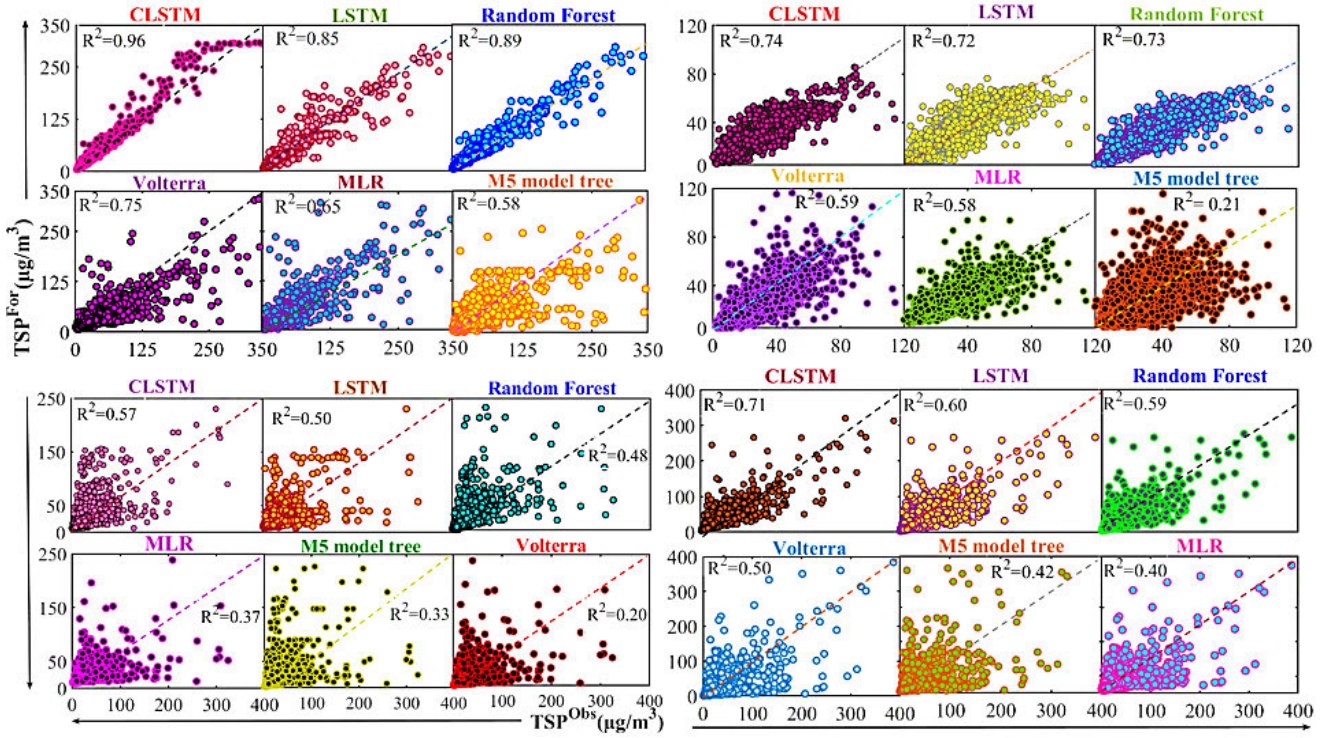

FIGURE 6. Scatterplot of forecasted ('For') *vs.* observed ('Obs') *TSP* in the model test phase. (a) Brisbane, (b) Townsville, (c) Hopeland, (d) Miles Airport. Note that the coefficient of determination (R^2) is indicated in each subplot.

the *CLSTM* model's accuracy is achieved through Pearson's correlation, 'r' comparing the hourly *TSP* (forecasted ('*For*') *vs.* observed ('*Obs*')) and the goodness-of-fit displayed *via* scatterplots (Fig. 6 (a)-(d)). Here, scatterplots are used to authenticate the consensus between regression, coefficient of determination, and predictors variables with a linear fit including the r^2 value are used to outline the model's accuracy [50]. Notably, for short-term forecasts, *CLSTM* gives the best performance to further ascertain its suitability. This deduction is supported by a high r^2-value for Brisbane (*CLSTM* $\approx$ *0.96*) *vs.* (*LSTM* $\approx$ *0.85*, Volterra $\approx$ *0.75*, *RF* $\approx$ *0.89*, M5 model tree $\approx$ *0.58*, and *MLR* $\approx$ *0.75*). Similarly, for Townsville and Miles Airport, the r^2–value for hybrid *CLSTM* is $\approx$ (*0.74, 0.71*) *vs. LSTM* $\approx$ (*0.72, 0.60*), *RF* $\approx$ (*0.73, 0.59*), Volterra $\approx$ (*0.59, 0.50*), *MLR* $\approx$ (*0.58, 0.40*), and M5 model tree $\approx$ (*0.21, 0.42*). It is noticed that, although all models performed relatively poorly for the case of Hopeland, the performance of *CLSTM* was better ($\approx$ *0.57*) than the rest of the models. Importantly, the standalone models produced inferior results in forecasting hourly *TSP* with r^2 values of *LSTM* ($\approx$ *0.50*), *M5* ($\approx$ *0.33*), *RF* ($\approx$ *0.48*), *MLR* ($\approx$ *0.37*), and Volterra ($\approx$ *0.20*). These results concur with the high values of *MAPE* and *RMSE* for this station as elaborated in further discussion. The proposed hybrid *CLSTM* model also showed superior performance *w.r.t* mean absolute error (*MAE)*. For instance, in the case of Brisbane, *MAE* $\approx$ *3.54*μg/m^3 as compared to *MAE* $\approx$ *4.69*μg/m^3 (*LSTM*), *MAE* (*RF* $\approx$ *15.51*μg/m^3, Volterra $\approx$ *7.91*μg/m^3, *MLR* $\approx$ *8.13*μg/m^3, M5 model tree $\approx$ *11.66*μg/m^3). Importantly, *MAE* is significantly lower for all stations as compared to the remaining ensemble of models in the research.

It concurs from Table 5(a)–(d), that as the magnitude of *MAE* for the proposed hybrid *CLSTM* model is low ($<$ 10%) for all the study sites, this generally falls in the category of an excellent model, see Ref [60]. However, there also appears to be a subtle variation in model accuracy when the root mean square error (*RMSE*) and the mean absolute percentage error (*MAPE)* is computed. For instance, the hybrid *CLSTM* model is seen to generate an *RMSE* $\approx$ *13.97*μg/m^3, *7.44*μg/m^3, and *MAPE* $\approx$ *8.43%*, *MAPE* $\approx$ *17.37%* for Brisbane and Townsville; (*i.e.*, a good model category with an error $<$ 20%), and fair results with *RMSE* $\approx$ *21.97*μg/m^3, *23.11*μg/m^3 for Hopeland and Miles Airport (*i.e.*, a fair model category with error $>$ 20%). Notably, the two stations (Hopeland & Miles Airport) generally incur high *RMSE* and *MAPE*, however, the lowest *RMSE* is registered by the hybrid *CLSTM* model in such a way that the error for the other models occurs in a higher range *i.e.*, [*24.23– 29.62*] μg/m^3 for Hopeland. This is also noted for the case of Miles Airport that has the lowest error for the hybrid *CLSTM* model, while the other models are seen to obtain a higher error in the range [*26.66– 49.35*] μg/m^3. Following our results presented to far, we can infer that *CLSTM* accedes to much greater accuracy than the five other models, and thus appears to be a versatile predictive tool in modelling *TSP* at hourly steps. To further confirm this, we revert to radial (or

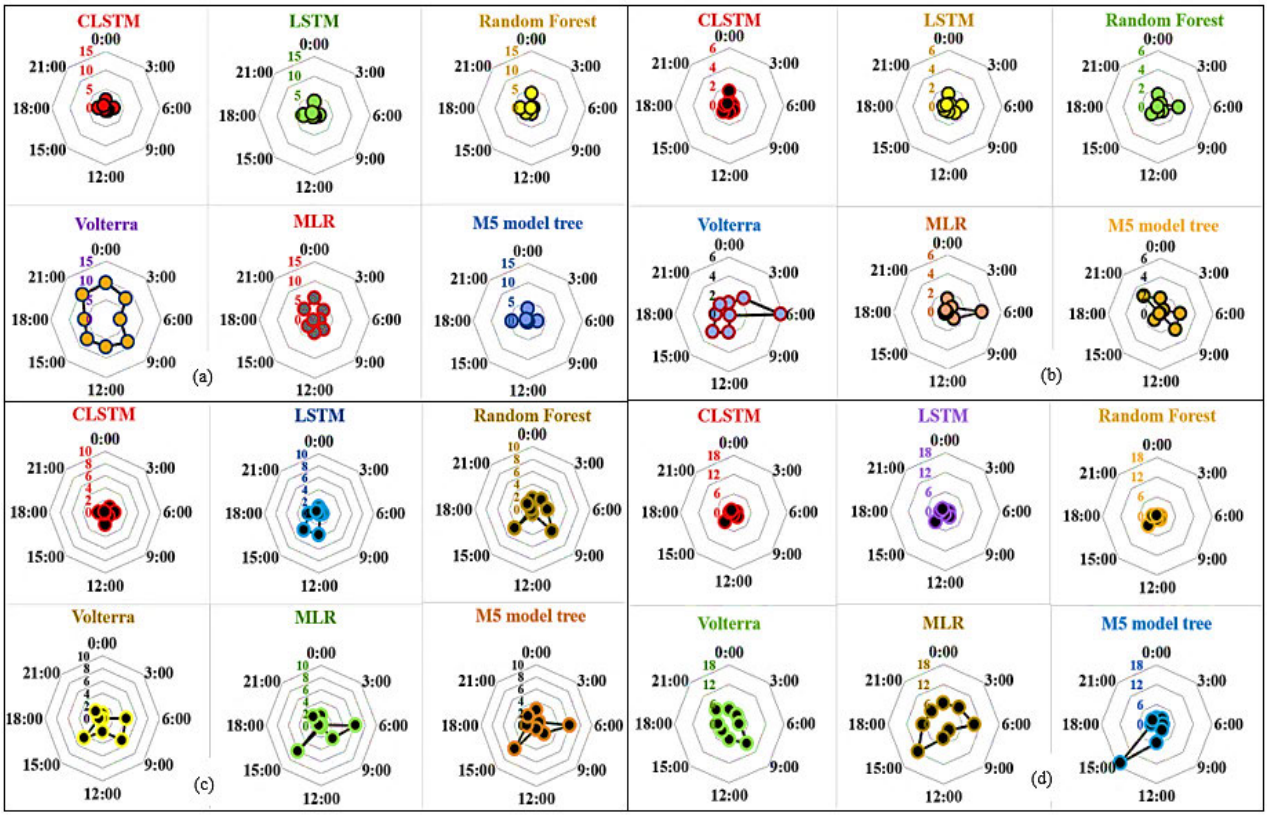

FIGURE 7. Relative forecasted error (%) in *TSP* generated by *CLSTM* vs. five other competing models. (a) Brisbane, (b) Townsville, (c) Hopeland, (d) Miles Airport sites. Note that the axis from origin denotes the proportion of model error per hour.

polar) plots showing the relative (or percentage) forecast error in Fig. 7 (a)-(d). In principle, the performance can be denoted as good when forecasted error (*FE*) is close to trivial. Therefore, the hybrid *CLSTM* model's results are seen to excel to a much greater degree in comparison to its counterpart models. For all study sites, these show an *FE* (%) value from midnight (0.00–23.00) in such a way that the *CLSTM* model elucidates a high degree of accuracy. This performance evaluation also shows that *CLSTM* was succeeded by *RF*, Volterra, *MLR,* and M5 tree with the latter showing high relative *FE* computed at an hourly time-step. Fig. 8 (a)-(d) illustrates a comparison of forecasted *vs.* observed *TSP* where data are averaged for the entire test phase. We have used red to show *CLSTM* simulations and black for observed data. The plots illustrate the preciseness of the objective model against competing approaches. In closing, Fig. 9 (a) illustrates the forecasted the testing ('*For*') *vs.* observed ('*Obs*') *TSP* at the hourly interval in phase for one of the study sites (Townsville). The hybrid *CLSTM* model is seen to yield comparatively high accuracy ($r \approx 0.83$, $MAE \approx 5.40 \mu g/m^3$, $RMSE \approx 7.44 \mu g/m^3$, $WI \approx 0.91$) between TSP^{Obs} and TSP^{For}. Importantly, the forecasted and observed *TSP* are quite adjacent to each other, suggesting a potential benefit of using *CLSTM* for reliable *AQ* forecasting systems. Fig. 9 (b-d) further shows the efficacy of the hybrid *CLSTM* model illustrating the model *MAE* loss and loss comparison in the training and validation phase from January 2015 – June 2018 for the Townsville study site. To summarise, we aver that the newly proposed deep learning *CLSTM* model has performed very well for the study sites *w.r.t* stringent error evaluation such as, but not limited to the Willmott's Index of agreement, Nash–Sutcliffe Efficiency and Legates and McCabe Index. This was supported by a high correlation of forecasted and observed *TSP* as well as the lowest errors. Following our results, we state that the use of a non-*DL* approach (*e.g.*, *RF*, *M5 Tree*, or *MLR*) is likely to result in inferior performance compared with a hybrid *DL* model (*e.g.*, *CLSTM*) especially for data such as *TSP* that can be highly chaotic in their behaviours over short-term (*e.g.*, hourly) scale. Having said that, we note that the only study site where the performance of the *DL* models was not excellent; but still better than its counterpart models, was Hopeland. On closer observation, we see that this site has a high elevation (316m), as shown in Table 1, and while the rest of the pollutants ($PM_{2.5}$ and PM_{10}) are not considered in its model building, greenhouse gases and solar radiation were considered apart from the other model input variables. This is shown in Fig. 3 and Table 3, which might be a factor affecting the performance of the *DL* model. Similarly, the site where the performance of the *DL* model was best, compared to all others, was Brisbane. On closer observation, we note

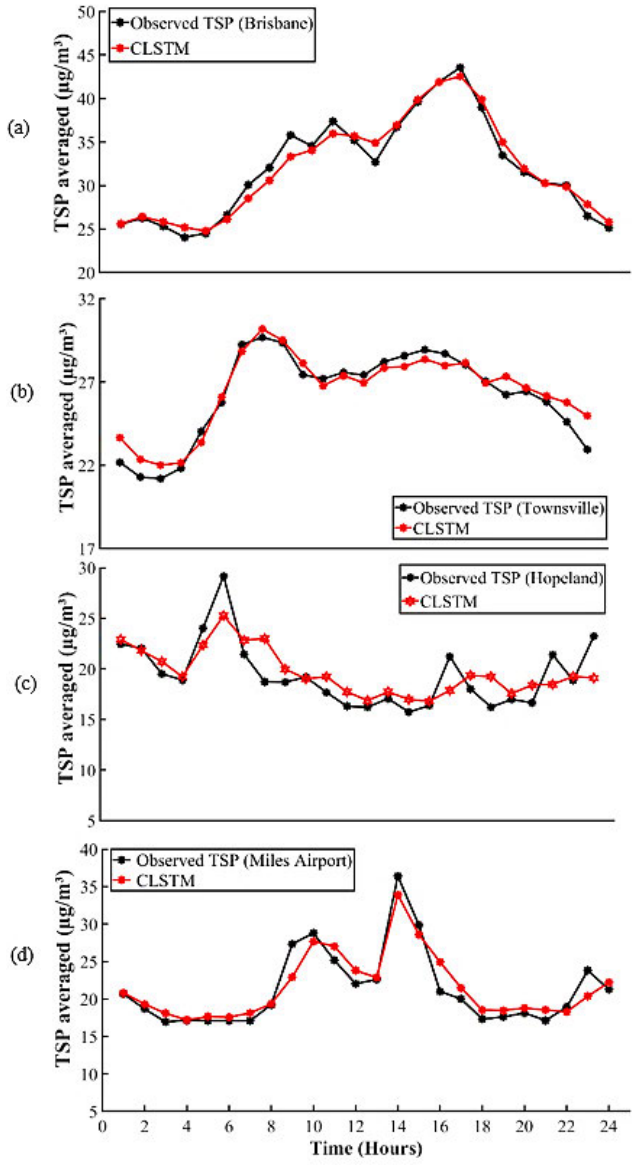

FIGURE 8. Forecasted *vs.* observed *TSP* in the test phase. (a) Brisbane, (b) Townsville, (c) Hopeland, (d) Miles Airport.

this site has the lowest elevation (9m; Table 1, Table 3) where none of the gases and solar radiation were considered apart from other selected model input variables. The difference in the availability of model creation data might therefore be a pertinent factor causing a range of differences in the performance of these models.

In summary, the newly proposed deep learning hybrid (*i.e.*, *CLSTM*) model provided a relatively precise performance *i.e.* highest (r), highest (E_{NS}), including a large magnitude of the most stringent model performance metric *i.e.* L, including the lowest *RMSE, MAPE,* and *MAE* compared with the five other benchmark models. Consequently, these results provided compelling evidence, establishing the deep learning hybrid *CLSTM* model as a credible and relatively practical framework for *TSP* predictions. The intelligent framework for modelling air pollutants can also have a major application in the academic and industrial settings, as this particulate matter poses a major health issue. We thus conclude that *CLSTM* can be considered as a novel framework for pollutant modelling, and therefore, can be used as a pragmatic tool in monitoring the atmospheric environment.

V. CONCLUSION

Using deep learning approach, this study reports the potential utility of an air pollution forecasting system developed and evaluated at hourly timesteps. The research work has designed a hybrid predictive model (*i.e.*, *CLSTM*) adopted to forecast the total suspended particulate matter. The newly proposed *CLSTM* model amalgamated convolutional neural network with long short-term memory network to attain optimal performance. For enhanced performance utility, our *CLSTM* firstly employed the *CNN* algorithm to automatically detect and extract important features from predictor variables while *LSTM* collated these features to generate a time series for the next modelling phase. Elucidated by charts, plots, and statistical metrics of forecasted and observed *TSP*, the findings reveal a superior performance of *CLSTM* relative to an ensemble of five competing models. Our findings, including the contributions of this study, are as follows:

i) The paper has bridged significant gaps in modelling hourly *TSP* for an *APF* system, especially for Australia by presenting a constructive research methodology to generate a computationally efficient architecture denoted as a hybrid *'CLSTM'* model.

ii) The ensemble of five other competing models: Random Forest, Volterra, *MLR*, M5 model tree, and *LSTM* appeared to lag in their capability to generate satisfactory forecasts in comparison to the deep learning *CLSTM* hybrid model. Mean absolute and root mean square error for all competing models were significantly higher than the deep learning hybrid model. For argument sake, our simulations registered *MAE* (μg/m^3) = (3.54 against *4.69, 15.51, 7.91, 8.13, 11.66*), and *RMSE* value (μg/m^3) = (13.97 *vs. 23.31, 40.74, 19.24, 20.27, 46.71*) for *CLSTM vs. LSTM*, *RF*, Volterra, *MLR* or M5 model tree algorithms, respectively.

iii) A comprehensive evaluation with statistical metrics, charts, and plots of tested data demonstrated that the hybrid *CLSTM* model was able to yield relatively better forecasts in comparison with the five comparative benchmark models. The performance metrics illustrated the *CLSTM* model to attain efficient performance for short-term *TSP* forecasts. For example, the most rigorous statistical metrics, L, was superior for all four stations with values of (*0.80, 0.51, 0.35, 0.61*) for *CLSTM vs.* (*0.73, 0.45, 0.35, 0.53*) for *LSTM,* (*0.64, 0.44,0.29, 0.46*) for *RF*, (*0.54, 0.42, 0.19, 0.44*) for Volterra, (*0.53, 0.43, 0.27, 0.28*) for *MLR,* and (*0.33, 0.36, 0.20, 0.17*) for M5 model tree. These metrics are providing undisputed evidence that the *CLSTM*

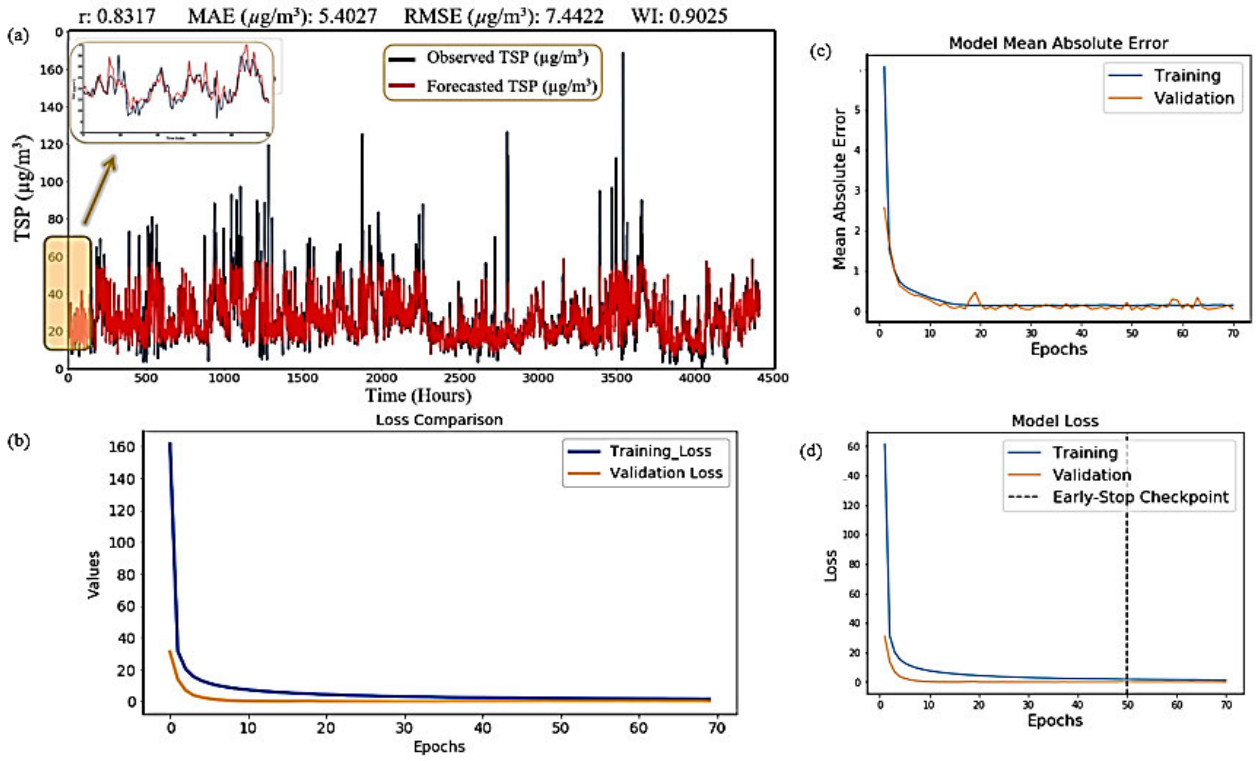

FIGURE 9. (a-d).Performance evaluation with epoch = 2000 at Townsville in the test phase (July-December 2018).

model was very successful in forecasting the hourly air pollutant datasets for all sites.

iv) The percentage error for *CLSTM* also suggested that the fusion of *CNN* and *LSTM* frameworks led to better predictive outcomes, as verified by *MAPE* (*8.43% w.r.t CLSTM vs. 10.20%* (*LSTM*), *22.01%* (*RF*), *25.12%* (Volterra), *27.08%* (*MLR*), and *36.07%* (M5 model tree). In both the training and the testing phase the proposed *CLSTM* hybrid framework was in the 'excellent model category, see Ref [60], as exemplified by *MAE* $< 10\%$ (at all study sites), with low *RMSE* $< 10\%$ for Townsville or $< 15\%$ for Brisbane. This reveals that *CLSTM* architecture was considerably robust in forecasting hourly data, and therefore indicated its possible application in the air pollution model and future public health and air quality studies.

VI. LIMITATIONS AND FUTURE RESEARCH

Despite the efficacy of *CLSTM*, there remain some limitations that can be used to recommend future research.

- Independent testing of the *CLSTM* model at different time steps such as at short-term (*e.g.*, 1–, 5– or 10–minute) resolution may enhance its practical adoption from an end-user perspective. The very short-term forecasting can also lead to a more proactive application of *CLSTM*, in terms of public advisory roles, and especially, the vulnerable populations *e.g.*, children, people with medical issues, senior citizens, and expectant women and planning accordingly.
- In future research, one can improve the *CLSTM* model using ancillary tools such as bidirectional *LSTM* algorithms to generate better outcomes. Also, considering the chaotic nature of air pollutants, the use of an improved complete ensemble empirical mode decomposition with an adaptive noise algorithm to extract temporal information of air pollutant movements could further add value to the proposed *APF* system.
- It is also recommended that future research could incorporate the Internet of Things (*IoT*) to convert the hybrid *CLSTM* model into an application (*app*) for the Internet and mobile phone users. This could be similar to a recently proposed system for forecasting solar ultraviolet (*UV*) radiation tailored to generate very short-term reactive *UV* forecasts for public health risk mitigation [62]. Similarly, an IoT system for air quality forecasting can increase the practicality of *CLSTM*, particularly, ensuring greater public use of the newly proposed *AQ* system.

In closing, the study avers that the deep learning hybrid model: *CLSTM* model carries significant merits, especially in atmospheric modelling effort, assisting in the evaluation of the impact of pollutants on health and environment and contributing to research on air quality and general risk assessments. Consequently, this research work is a paramount step regarding the construction of an effective air pollutant forecasting technology making a vital impact on the environment and health risk alleviation.

ACKNOWLEDGMENT

The authors are thankful to the reviewers for their helpful suggestions and the Queensland Department of Environment and Science for the data.

REFERENCES

[1] P. Gupta, S. A. Christopher, M. A. Box, and G. P. Box, "Multi year satellite remote sensing of particulate matter air quality over Sydney, Australia," *Int. J. Remote Sens.*, vol. 28, no. 20, pp. 4483–4498, Oct. 2007.

[2] M. Oprea, S. F. Mihalache, and M. Popescu, "A comparative study of computational intelligence techniques applied to PM2.5 air pollution forecasting," in *Proc. 6th Int. Conf. Comput. Commun. Control (ICCCC)*, May 2016, pp. 103–108.

[3] X. Xu, H. Ding, and X. Wang, "Acute effects of total suspended particles and sulfur dioxides on preterm delivery: A community-based cohort study," *Arch. Environ. Health, Int. J.*, vol. 50, no. 6, pp. 407–415, Dec. 1995.

[4] E. Sharma, R. C. Deo, R. Prasad, and A. V. Parisi, "A hybrid air quality early-warning framework: An hourly forecasting model with online sequential extreme learning machines and empirical mode decomposition algorithms," *Sci. Total Environ.*, vol. 709, Mar. 2020, Art. no. 135934.

[5] M. R. Heal, P. Kumar, and R. M. Harrison, "Particles, air quality, policy and health," *Chem. Soc. Rev.*, vol. 41, no. 19, pp. 6606–6630, 2012.

[6] J. Kukkonen *et al.*, "A review of operational, regional-scale, chemical weather forecasting models in Europe," *Atmos. Chem. Phys.*, vol. 12, no. 1, pp. 1–87, Jan. 2012.

[7] K. D. Karatzas and S. Kaltsatos, "Air pollution modelling with the aid of computational intelligence methods in Thessaloniki, Greece," *Simul. Model. Pract. Theory*, vol. 15, no. 10, pp. 1310–1319, Nov. 2007.

[8] S. Pan, Y. Choi, A. Roy, and W. Jeon, "Allocating emissions to 4 km and 1 km horizontal spatial resolutions and its impact on simulated NO_x and O_3 in Houston, TX," *Atmos. Environ.*, vol. 164, pp. 398–415, Sep. 2017.

[9] S. Ghimire, R. C. Deo, N. Raj, and J. Mi, "Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms," *Appl. Energy*, vol. 253, Nov. 2019, Art. no. 113541.

[10] C.-Y. Zhang, C. L. P. Chen, M. Gan, and L. Chen, "Predictive deep Boltzmann machine for multiperiod wind speed forecasting," *IEEE Trans. Sustain. Energy*, vol. 6, no. 4, pp. 1416–1425, Oct. 2015.

[11] S. Ghimire, R. C. Deo, N. Raj, and J. Mi, "Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms," *Appl. Energy*, vol. 253, Nov. 2019, Art. no. 113541.

[12] S. Ghimire, R. C. Deo, N. Raj, and J. Mi, "Deep learning neural networks trained with MODIS satellite-derived predictors for long-term global solar radiation prediction," *Energies*, vol. 12, no. 12, p. 2407, Jun. 2019.

[13] G. Bargshady, X. Zhou, R. C. Deo, J. Soar, F. Whittaker, and H. Wang, "Enhanced deep learning algorithm development to detect pain intensity from facial expression images," *Expert Syst. Appl.*, vol. 149, Jul. 2020, Art. no. 113305.

[14] H. Al-Hadeethi, S. Abdulla, M. Diykh, R. C. Deo, and J. H. Green, "Adaptive boost LS-SVM classification approach for time-series signal classification in epileptic seizure diagnosis applications," *Expert Syst. Appl.*, vol. 161, Dec. 2020, Art. no. 113676.

[15] E. Eslami, Y. Choi, Y. Lops, and A. Sayeed, "A real-time hourly ozone prediction system using deep convolutional neural network," *Neural Comput. Appl.*, vol. 32, no. 13, pp. 8783–8797, Jul. 2020.

[16] X. Li, L. Peng, X. Yao, S. Cui, Y. Hu, C. You, and T. Chi, "Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation," *Environ. Pollut.*, vol. 231, pp. 997–1004, Dec. 2017.

[17] Q. Zhang, J. C. K. Lam, V. O. K. Li, and Y. Han, "Deep-AIR: A hybrid CNN-LSTM framework forFine-grained air pollution forecast," 2020, *arXiv:2001.11957*. [Online]. Available: http://arxiv.org/abs/2001.11957

[18] T. Li, M. Hua, and X. Wu, "A hybrid CNN-LSTM model for forecasting particulate matter (PM2.5)," *IEEE Access*, vol. 8, pp. 26933–26940, 2020.

[19] G. P. V. Athira, R. Vinayakumar, and K. Soman, "DeepAirNet: Applying recurrent networks for air quality prediction," in *Proc. Int. Conf. Comput. Intell. Data Sci.* vol. 132, 2018, pp. 1394–1403.

[20] J. Chen, W. Quan, T. Cui, and Q. Song, "Estimation of total suspended matter concentration from MODIS data using a neural network model in the China eastern coastal zone," *Estuarine, Coastal Shelf Sci.*, vol. 155, pp. 104–113, Mar. 2015.

[21] L.-S. Wen, C.-P. Lee, W.-H. Lee, and A. Chuang, "An ultra-clean multilayer apparatus for collecting size fractionated marine plankton and suspended particles," *J. Visualized Exp.*, no. 134, p. e56811, Apr. 2018, doi: 10.3791/56811.

[22] Q. Guo, Z. Zhu, Z. Cheng, S. Xu, X. Wang, and Y. Duan, "Correction of light scattering-based total suspended particulate measurements through machine learning," *Atmosphere*, vol. 11, no. 2, p. 139, Jan. 2020.

[23] N. Wang, R. Liu, N. Asmare, C.-H. Chu, and A. F. Sarioglu, "Processing code-multiplexed coulter signals via deep convolutional neural networks," *Lab Chip*, vol. 19, no. 19, pp. 3292–3304, 2019.

[24] G. Liu and S. Shuo, "Air quality forecasting using convolutional LSTM," Stanford Univ., Stanford, CA, USA, Tech. Rep. 8291197, Spring 2018. [Online]. Available: http://cs230.stanford.edu/projects_spring_2018/reports/8291197.pdf

[25] G. Bargshady, X. Zhou, R. C. Deo, J. Soar, F. Whittaker, and H. Wang, "Enhanced deep learning algorithm development to detect pain intensity from facial expression images," *Expert Syst. Appl.*, vol. 149, Jul. 2020, Art. no. 113305.

[26] C. Mathers, T. Vos, and C. Stevenson, "The burden of disease and injury in Australia," *Austral. Health Rev.*, vol. 23, no. 1, p. 216, 2000.

[27] S. Davey and A. Sarre, *The 2019/20 Black Summer Bushfires*. New York, NY, USA: Taylor & Francis, 2020.

[28] P. Yu, R. Xu, M. J. Abramson, S. Li, and Y. Guo, "Bushfires in Australia: A serious health emergency under climate change," *Lancet Planet. Health*, vol. 4, no. 1, pp. e7–e8, Jan. 2020.

[29] M. Hendryx, N. Higginbotham, B. Ewald, and L. H. Connor, "Air quality in association with rural coal mining and combustion in new south Wales Australia," *J. Rural Health*, vol. 35, no. 4, pp. 518–527, Sep. 2019.

[30] R. Prasad, R. C. Deo, Y. Li, and T. Maraseni, "Ensemble committee-based data intelligent approach for generating soil moisture forecasts with multivariate hydro-meteorological predictors," *Soil Tillage Res.*, vol. 181, pp. 63–81, Sep. 2018.

[31] M. Ali, R. Prasad, Y. Xiang, and R. C. Deo, "Near real-time significant wave height forecasting with hybridized multiple linear regression algorithms," *Renew. Sustain. Energy Rev.*, vol. 132, Oct. 2020, Art. no. 110003.

[32] A. Rahimikhoob, M. Asadi, and M. Mashal, "A comparison between conventional and m5 model tree methods for converting pan evaporation to reference evapotranspiration for semi-arid region," *Water Resour. Manage.*, vol. 27, no. 14, pp. 4815–4826, Nov. 2013.

[33] A. Suleiman, M. R. Tight, and A. D. Quinn, "A comparative study of using random forests (RF), extreme learning machine (ELM) and deep learning (DL) algorithms in modelling roadside particulate matter (PM10 & PM2.5)," in *Proc. IOP Conf. Ser., Earth Environ. Sci.*, 2020, Art. no. 012126.

[34] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.

[35] A. Alléon, G. Jauvion, B. Quennehen, and D. Lissmyr, "PlumeNet: Large-scale air quality forecasting using a convolutional LSTM network," 2020, *arXiv:2006.09204*. [Online]. Available: http://arxiv.org/abs/2006.09204

[36] C.-J. Huang and P.-H. Kuo, "A deep CNN-LSTM model for particulate matter (PM2.5) forecasting in smart cities," *Sensors*, vol. 18, no. 7, p. 2220, Jul. 2018.

[37] S. Oehmcke, O. Zielinski, and O. Kramer, "Input quality aware convolutional LSTM networks for virtual marine sensors," *Neurocomputing*, vol. 275, pp. 2603–2615, Jan. 2018.

[38] S. Du, T. Li, Y. Yang, and S.-J. Horng, "Deep air quality forecasting using hybrid deep learning framework," *IEEE Trans. Knowl. Data Eng.*, early access, Nov. 20, 2019, doi: 10.1109/TKDE.2019.2954510.

[39] J. Chen, G.-Q. Zeng, W. Zhou, W. Du, and K.-D. Lu, "Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization," *Energy Convers. Manage.*, vol. 165, pp. 681–695, Jun. 2018.

[40] M. Sundermeyer, R. Schlüter, and H. Ney, "LSTM neural networks for language modeling," Tech. Rep., 2012.

[41] S. Han, Y. Wang, H. Yang, W. J. Dally, J. Kang, H. Mao, Y. Hu, X. Li, Y. Li, D. Xie, H. Luo, and S. Yao, "ESE: Efficient speech recognition engine with sparse LSTM on FPGA," in *Proc. ACM/SIGDA Int. Symp. Field-Program. Gate Arrays (FPGA)*, 2017, pp. 75–84.

[42] Y.-T. Tsai, Y.-R. Zeng, and Y.-S. Chang, "Air pollution forecasting using RNN with LSTM," in *Proc. IEEE 16th Int. Conf. Depend., Autonomic Secure Comput., 16th Int. Conf Pervas. Intell. Comput., 4th Int. Conf Big Data Intell. Comput. Cyber Sci. Technol. Congr. (DASC/PiCom/DataCom/CyberSciTech)*, Aug. 2018, pp. 1074–1079.

[43] *Quarterly Update of Australia's National Greenhouse gas Inventory: Mar. 2018, Incorporating Emissions from the NEM up to June 2018*, CWA, Mar. 2018.

[44] *National Pollution Inventory, Department of Agriculture, Water and the Environment, CSIRO Submission no 48 to Senate Community Affairs References Committee, Parliament of Australia, Impacts on Health of Air Quality in Australia*, p. 8. [Online]. Available: http://www.npi.gov.au/resource/particulate-matter-pm10-and-pm25

[45] W. L. Junger and A. P. de Leon, "Imputation of missing data in time series for air pollutants," *Atmos. Environ.*, vol. 102, pp. 96–104, Feb. 2015.
[46] N. Ketkar, "Introduction to Keras," *Deep Learning With Python*. Cham, Switzerland: Springer, 2017, pp. 97–111.
[47] M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning," Tech. Rep., 2016, pp. 265–283.
[48] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.
[49] S. Sun and J.-L. Bertrand-Krajewski, "Input variable selection and calibration data selection for storm water quality regression models," *Water Sci. Technol.*, vol. 68, no. 1, pp. 50–58, Jul. 2013.
[50] R. C. Deo, X. Wen, and F. Qi, "A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset," *Appl. Energy*, vol. 168, pp. 568–593, Apr. 2016.
[51] R. C. Deo, M. K. Tiwari, J. F. Adamowski, and J. M. Quilty, "Forecasting effective drought index using a wavelet extreme learning machine (W-ELM) model," *Stochastic Environ. Res. Risk Assessment*, vol. 31, no. 5, pp. 1211–1240, Jul. 2017.
[52] R. C. Deo, M. K. Tiwari, J. F. Adamowski, and J. M. Quilty, "Forecasting effective drought index using a wavelet extreme learning machine (W-ELM) model," *Stochastic Environ. Res. Risk Assessment*, vol. 31, no. 5, pp. 1211–1240, Jul. 2017.
[53] N. Karunanithi, W. J. Grenney, D. Whitley, and K. Bovee, "Neural networks for river flow prediction," *J. Comput. Civil Eng.*, vol. 8, no. 2, pp. 201–220, Apr. 1994.
[54] M. Ali and R. Prasad, "Significant wave height forecasting via an extreme learning machine model integrated with improved complete ensemble empirical mode decomposition," *Renew. Sustain. Energy Rev.*, vol. 104, pp. 281–295, Apr. 2019.
[55] R. Yu, J. Gao, M. Yu, W. Lu, T. Xu, M. Zhao, J. Zhang, R. Zhang, and Z. Zhang, "LSTM-EFG for wind power forecasting based on sequential correlation features," *Future Gener. Comput. Syst.*, vol. 93, pp. 33–42, Apr. 2019.
[56] D. Zeng, "Workshop 3: Deep learning with Python," in *Proc. South Dakota State Univ. Symp. Session 8, Workshop 3, United States*, 2019. [Online]. Available: https://openprairie.sdstate.edu/datascience_symposium/2019/sessions/
[57] J. E. Nash and J. V. Sutcliffe, "River flow forecasting through conceptual models part I—A discussion of principles," *J. Hydrol.*, vol. 10, no. 3, pp. 282–290, Apr. 1970.
[58] D. R. Legates and G. J. McCabe, Jr., "Evaluating the use of 'goodness-of-fit' measures in hydrologic and hydroclimatic model validation," *Water Resour. Res.*, vol. 35, no. 1, pp. 233–241, 1999.
[59] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature," *Geoscientific Model Develop.*, vol. 7, no. 3, pp. 1247–1250, Jun. 2014.
[60] K. Mohammadi, S. Shamshirband, C. W. Tong, M. Arif, D. Petković, and S. Ch, "A new hybrid support vector machine–wavelet transform approach for estimation of horizontal global solar radiation," *Energy Convers. Manage.*, vol. 92, pp. 162–171, Mar. 2015.
[61] D. R. Legates and G. J. McCabe, "A refined index of model performance: A rejoinder," *Int. J. Climatol.*, vol. 33, no. 4, pp. 1053–1056, Mar. 2013.
[62] R. C. Deo, N. Downs, A. V. Parisi, J. F. Adamowski, and J. M. Quilty, "Very short-term reactive forecasting of the solar ultraviolet index using an extreme learning machine integrated with the solar zenith angle," *Environ. Res.*, vol. 155, pp. 141–166, May 2017.

EKTA SHARMA (Member, IEEE) received the B.Sc. degree in mathematical sciences, the M.Sc. degree in operational research, and the M.Phil. degree from the University of Delhi, India. She is currently pursuing the Ph.D. degree in artificial intelligence with the University of Southern Queensland, Australia. She has served in both academia and industry across varied roles as an Area Manager to a Learning Advisor, Lecturer, and Researcher with universities in Europe, India, and Australia. Her research work has received funding from the Australian Defence Science and Technology Group, The Australian Mathematical Sciences Institute, and the Australian Government. Her research interests include atmospheric sciences, environmental management, data analytics, and deep learning. She is a member of INFORMS and AMSI.

RAVINESH C. DEO (Senior Member, IEEE) is currently an Associate Professor with the University of Southern Queensland, Australia, with research interests in data science, decision systems, and deep learning. He published over 215 articles, including 160 journals (mostly Scopus Quartile 1), seven books under Elsevier, Springer Nature, and IGI Global, and ten book chapters. His work has cumulative citations exceeding 5,400 with H-index of 40. He has supervised over 18 research degrees; he was an Industry Mentor through externally funded grants and a visiting professor internationally and professionally, and is affiliated under scientific bodies. He was awarded internationally competitive fellowships, such as the Queensland *US* Smithsonian, Australia–China Scientist, *JSPS*, Chinese Academy of Science, Australia-India, and Endeavour Fellowship. He serves on Editorial Boards of *Stochastic Environmental Research and Risk Assessment*, *Remote Sensing* and *Energies* journals, including *Journal of Hydrologic Engineering* (ASCE), contributing to teaching, research, and service in data science, mathematics, and modeling.

RAMENDRA PRASAD received the B.Sc. degree in mathematics and physics and the M.Sc. degree in physics from The University of the South Pacific, Fiji, and the Ph.D. degree in modeling and simulations from the University of Southern Queensland, Australia, in 2019. Since 2018, he has been a Lecturer with The University of Fiji. He has authored 15 articles in peer-reviewed journals and reputed conferences. His research interests include advanced machine learning approaches, data analytics, hydrological modeling, energy modeling, energy management, environmental, atmospheric modeling, and ocean wave modeling.

ALFIO V. PARISI the B.Sc. degree (Hons.) in 1981, the M.App.Sc. degree in 1992, and the Ph.D. degree in 1996. He is currently a Professor of Physics with the Faculty of Health, Engineering, and Sciences, University of Southern Queensland, Australia. His research interests include solar UV dosimetry, radiometry, and spectroradiometry, including the development of new techniques on solar measurement, applied to provide improved characterization of solar UV environment, and measurement of solar exposures to plants.

NAWIN RAJ received the B.Sc., B.Ed., PGDMA, and M.Sc. degrees in computational fluid dynamics from The University of the South Pacific and the Ph.D. degree from the University of Southern Queensland (USQ), Australia, in 2015. From 2007 to 2010, he was a Lecturer with Fiji National University. He is currently a Lecturer at USQ. His research interests include artificial intelligence, deep learning, non-linear oscillation, computational fluid dynamics, and oceanography. He is a member of the Australian Mathematical Society and Queensland College of Teachers.