Deep learning for construction emission monitoring with low-cost sensor network

Huynh A.D. Nguyen[1], Trung H. Le[1], Quang P. Ha[1] and Merched Azzi[2]

[1]Faculty of Engineering Technology and IT, University of Technology Sydney, Australia,
[2]Department of Planning and Environment, New South Wales, Australia

{HuynhAnhDuy.Nguyen; Hoang.T.Le; Quang.Ha}@uts.edu.au,
Merched.Azzi@environment.nsw.gov.au

Abstract -
Emissions from construction activities, particularly in metropolitan areas, are carefully monitored to prevent health problems and environmental degradation. The data quality of low-cost wireless sensors in construction sites remains a challenge for pollution predictive models due to uncertainties of measurement and volatile environment. In this study, we propose a hybrid model using a Long short-term memory integrated with a Bayesian neural network to infer the probabilistic forecasts of particulate matters (i.e., $PM_{1.0}$, $PM_{2.5}$, and PM_{10}) emitted from construction activities. The training data are fused by two sources: (1) our developed low-cost wireless sensor network (LWSN) monitoring at a construction site located in Melrose Park, Sydney, Australia, and (2) air-quality stations (AQSs) in four suburbs nearby that monitoring site. The proposed model (LSTM-BNN) is compared with other deep learning methods, namely Gated recurrent unit (GRU), Bidirectional long short-term memory (BiLSTM) and One-dimension convolution neural network (1D-CNN), commonly used for time-series forecast. The experimental results indicate the outperformance of our model to all benchmark models and display a significant improvement at 56.3%, 27.9% and 37.9% in MAEs forecast for all three types of particles compared to a deterministic LSTM model.

Keywords -

Forecast; Time series; Deep Learning; Wireless sensor network; Dust monitoring

1 Introduction

Emission from construction sites is a common problem that adversely impacts on the health of workers, residents, and environment. The concentration levels of invisible particles (e.g., $PM_{1.0}$, $PM_{2.5}$ and PM_{10}) generated by a variety of construction activities, such as drilling, blasting, demolishing or earth moving need to be monitored and controlled [1, 2]. Recently, the development of IoT-enabled technology has improved the operational robustness of a low-cost wireless sensor network (LWSN) to stream reliable information in volatile conditions of the construction industry [3]. However, real-time data are insufficient for contractors to ensure compliance with governmental regulations over the environmental policy related to construction activities, of which noise and emissions often exceed the standards [4]. As such, an accurate forecast of the particle concentrations on the sites allows them to plan efficient activities and enable appropriate measures for minimizing dust emissions [5].

From a recent survey, despite the promising applications of machine learning (ML) and deep learning (DL) models in various tasks (e.g., site supervision, building inspection, safety detection and intelligent management [6]), there are only a few publications in predicting particles emitted from construction activities using on-site monitoring data. For example, a study from the UK proposed a deep neural network (DNN) in a scalable framework for highway air-quality monitoring [7], and in [8], the Bayesian optimization method was used to tune the hyperparameters of a long short term memory (LSTM) network trained by IoT-based data to predict $PM_{2.5}$ in subway construction. These studies used deterministic models of which the performance relies on only stationary data. Hence, to improve the prediction performance, a probabilistic approach appears promising to be explored for DL models in this area.

The estimation accuracy of DL models is determined by multiple components including but not limited to the modeling data, the choice of algorithm, model architectures and hyperparameters [9]. During the optimization process, intrinsic uncertainties of these components are required to be identified and mitigated by iterative training with updated data, especially to tackle a formidable task of forecasting particulate matter (PM) using LWSN data. Uncertainties involved in the process are aleatoric due to random noise in data as well as epistemic uncertainty arising from a lack of knowledge or information about the model or the process being modeled [9]. As such, addressing these uncertainties is crucial for creating reliable models that can make accurate predictions.

In fact, data gathered from LWSNs at construction sites are subject to inevitable noise, climatic factors and instability of physical systems [1, 10], resulting in significant aleatoric uncertainties. Additionally, on-site measurements are typically taken over a short period of time and often subject to information loss or missing, as determined by the project's duration and location. This limitation of data leads to epistemic uncertainty [9]. Therefore, a key question is how to handle uncertainties presented in LWSN data and model in order to improve the accuracy of DL algorithms for forecasting fine particles emitted from construction activities. This study proposes a DL model applying the long short term memory - recurrent neural network (LSTM-RNN) for modeling the time-series profiles of PM measured by LWSN fused with AQS data. An approximation of Bayesian neural network (BNN) for quantifying and mitigating uncertainties of model's inference is integrated with LSTM network to formulate a hybrid probabilistic model (LSTM-BNN) to improve the accuracy and reliability of the estimation.

The main contribution of this work rests with the capability of the proposed DL hybrid model in accurate and reliable prediction of emissions from construction sites in suburban areas by using the fusion of low-cost wireless sensor networks (LWSN) and air quality stations (AQS), by overcoming the two challenging factors: (1) limited amount of on-site data of air pollutants to train a forecast model and (2) reliability of the data collected due to random noise of the low-cost sensors.

The paper is organized as follows. After the Introduction, Section 2 presents the system framework including sequential steps from data collection to probabilistic inference of forecast values. Section 3 is devoted to model structures of LSTM and BNN approximation. The experimental results and benchmark with other deterministic DL models (i.e., LSTM, GRU, BiLSTM, and CNN) will be shown by important evaluation metrics in Section 4. Finally, a conclusion is drawn in Section 5.

2 System framework

The system framework is presented in Figure 1 comprising five sequential procedures, namely data collection, data processing, data preparation, model training and predictive inference.

2.1 Data collection

In this study, we used two sources of air-quality monitoring data to train and validate the performance of our proposed LSTM-BNN model. The first data source was from a developed LWSN implemented practically with a reliable sensing scheme, namely Wireless Dependable Sensing (W-DepS) networks published in our previous study [3]. This sensor network including 15 sensor motes was deployed at fixed locations within the construction site of Melrose Park in the state of New South Wales (NSW), Australia. Each device was equipped with sensors for measuring continuously air temperature, relative humidity and air-pollutant parameters such as particulate matters ($PM_{1.0}$, $PM_{2.5}$ and PM_{10}). The second data source was obtained from the AQS managed by the NSW government at four suburbs within a 10-km radius of the construction site, namely Parramatta North, Macquarie Park, Lidcombe and Rozelle [11]. Next, the crucial steps of processing data will be conducted before training model.

2.2 Data processing

Since the values from LWSN are asynchonized due to the latency of the wireless communication with the measured period of 15 minutes, the sensory data are firstly aligned in the same time stamp t for all values recorded within the interval of $(t - 7.5\ min, t + 7.5\ min]$ [3]. Besides, the hourly-averaging values of AQS are resampled and interpolated to the same 15-minute interval with the re-aligned sensory data. Next, the data noise and outliers are removed by a moving-average filter and the Cook's distance method expressed in the Equations 1 and 2:

$$x_{filter} = \frac{1}{n} \sum_{i=-\frac{n-1}{2}}^{\frac{n-1}{2}} x_i, \qquad (1)$$

where x_{filter} is the output of the filter, x_i is a range of values being averaged, and n is the range of centered samples (an odd number).

$$D_i = \frac{\sum_{j=1}^{n} (\hat{x}_j - \hat{x}_{(j)i})^2}{c.MSE}, \qquad (2)$$

where D_i is the Cook's distance of the observation i^{th}, x_i and $x_{(j)i}$ are respectively the fitted values when including and excluding samples i^{th}, MSE is the mean square error of two datasets, and c is the number of coefficients of the fitting model. Here, those samples were selected at least 3 times the means of the outliers' values [3].

Finally, the Min-Max normalization is applied to assure a synchronized scale for all variables. This step also contributes to faster optimization as the converging time of model parameters is reduced significantly during model training. The formula of normalizing data in a range of [0-1]:

$$x_{scaled} = \frac{x_{raw} - X_{min}}{X_{max} - X_{min}}, \qquad (3)$$

where x_{scaled} is a normalized value, x_{raw} is a raw value to be normalized, X_{max} and X_{min} are respectively the maximum and minimum values of the whole dataset.

2.3 Data preparation

Before training an RNN model in a supervision approach, we divide the time series D into sequences including pairs of an input sequence (X_i) and an output sequence (Y_i) which formulate a training sample $D_i(X_i, Y_i)$. These pairs of samples are split in a walk-forward fashion to attain the dynamics of time series ($D = \{D_1(X_1, Y_1), D_2(X_2, Y_2), ..., D_n(X_n, Y_n)\}$) as presented on the top right of Figure 1. In addition, this approach also enhances the number of samples for training the LSTM-BNN model presented in the next section.

Finally, the total processed data are approximately 15000 samples per variable over a period of 6 months

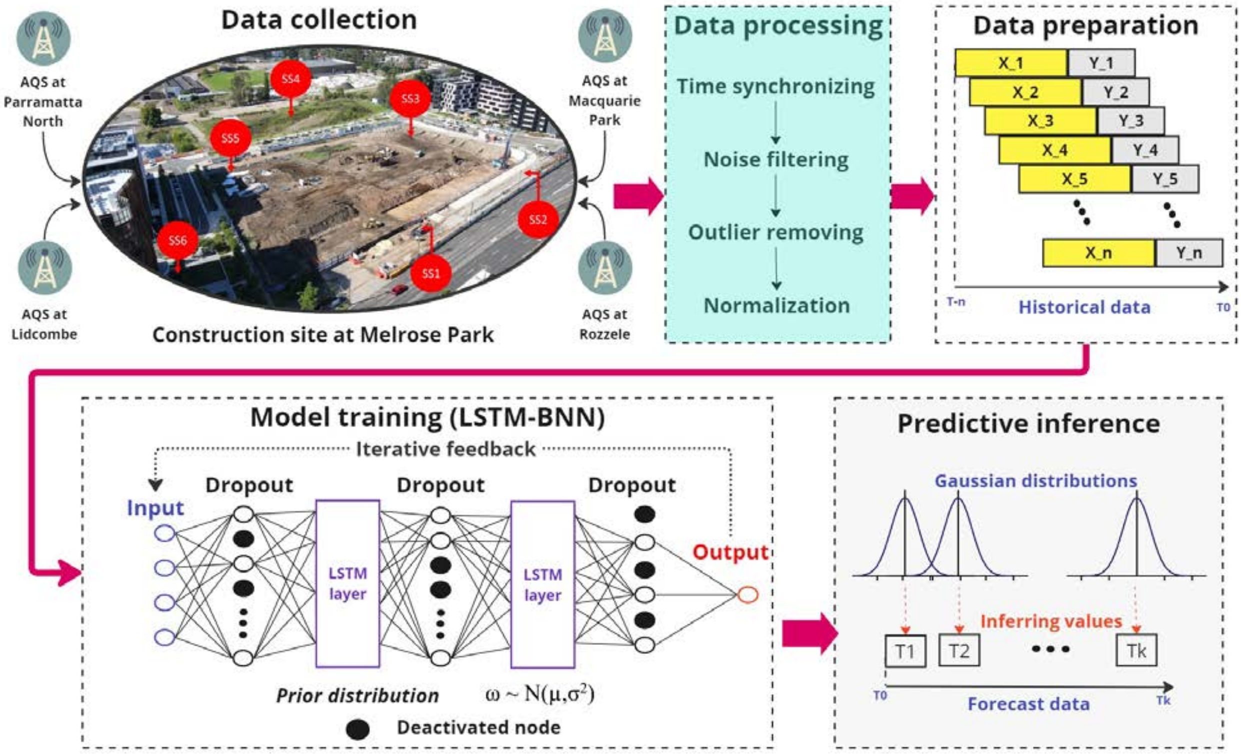

Figure 1. System framework for particles forecast with LSTM-BNN using LWSN (red nodes) located at a construction site at Melrose Park and AQS data from four nearby suburbs

(from November 2019 to May 2020). Here, three types of dust ($PM_{1.0}$, $PM_{2.5}$ and PM_{10}) measured by 15 low-cost sensors and two main particles ($PM_{2.5}$ and PM_{10}) from 4 neighbor AQSs are accounted for input variables.

During the monitoring period, there were two environmental incidents directly affecting the concentration levels in NSW, Australia: bushfires (Nov. 2019 - Feb. 2020) [12] and COVID-19 lockdown (from Mar. 2020) [3]. Hence, we partition 80% of the prepared data into the training set to ensure that the proposed model can capture the underlying patterns and generalize the data distributions over these events [8]. The remaining 20% of samples (i.e., equivalent to 36 days of observation) is divided equally into validating and testing sets to tune and evaluate the model performance.

3 Proposed learning model

3.1 LSTM-BNN network

The LSTM network is a variant of RNN proposed by Hochreiter and Schmidhuber in 1997 [13]. The key innovation of LSTM is the introduction of memory cells with three inputs of data x_t, the cell state C_{t-1}, and the hidden state h_{t-1} of the previous cell. As such, the long-term dependencies of time series are modeled well to adapt to periodic patterns and trends in air-quality forecast problems. The following equations express the flow of information into and out of the LSTM cells:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t, \tag{4}$$

where f_t and i_t are respectively the forget and input gates:

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f), \tag{5}$$

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i), \tag{6}$$

$$\tilde{C}_t = tanh(W_C.[h_{t-1}, x_t] + b_C). \tag{7}$$

The hidden state h_t is determined as a function of the cell state C_t:

$$h_t = o_t * tanh(C_t), \tag{8}$$

where the output gate o_t is determined as:

$$o_t = \sigma(W_f.[h_{t-1}, x_t] + b_o), \tag{9}$$

σ and $tanh$ represent respectively sigmoid and hyperbolic activation functions:

$$\sigma(x) = \frac{1}{1 + e^{-x}}, \tag{10}$$

$$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \tag{11}$$

The learnable parameters W_f, W_i, W_C, and b_f, b_i and b_C are respectively the weights and biases of the three gates. The cell state C_t is updated by an element-wise product ($*$) of the forget gate with the previous state ($f_t * C_{t-1}$) to skip the unimportant features and add up with the new feature from the input gate ($i_t * \tilde{C_t}$).

In this work, one input layer, two hidden LSTM layers (256 and 128 units), and one output layer formulate the proposed network with Rectified Linear Unit (ReLU) activation functions attached to both hidden layers [14].

$$ReLU(x) = \begin{cases} x & if\ x > 0, \\ 0 & otherwise. \end{cases} \tag{12}$$

The drop-out layers are inserted between the above-mentioned layers, taking the responsibility of a regularizer for reducing overfit problem [15]. This technique deactivates randomly the layer's nodes with a predefined probability p to force model learning multiple independent representations of the data, and hence it approximates the Gaussian process of BNN [16], and forms a hybrid network that we call the LSTM-BNN model.

Some other configurations for training the LSTM-BNN model include (i) Adam optimizer with mean square error metrics, (ii) the learning rate ($\lambda = 3e{-}4$), and (iii) an early-stopping function which terminates the learning process at a certain training iteration (epoch) when the model begins to overfit [17]. After training model, the next step will be inferring the prediction values. To quantify the uncertainties from data of LWSN and the model's configurations, an approximation of Bayesian inference is applied with the drop-out method to produce the predictive distributions at each forecast time step.

3.2 Bayesian inference approximation

In a BNN [18], the distribution of model's parameters ω are updated given the training data X_{train} following Bayesian theorem:

$$p(\omega|X_{train}) = \frac{p(X_{train}|\omega)p(\omega)}{p(X_{train})}, \tag{13}$$

where $p(X_{train}|\omega)$ is the likelihood of input values given ω with the prior $p(\omega)$, and $p(X_{train})$ is the marginal likelihood for the input distribution. The prior is initialized model's weights updated from the previous batches of data in each training epoch, sampled from parameters ω, assumed to follow the Gaussian distribution ($\omega \sim \mathcal{N}(\mu_\omega, \sigma_\omega)$).

When inferring outputs, the BNN model produces a predictive distribution from the weight distribution ω given a new input X_{new}:

$$p(y_{out|X_{new}}) = \int p(y_{out}|\omega)p(\omega|X_{new})d\omega. \tag{14}$$

Since the posterior of weights $p(\omega|X_{new})$ is intractable, its approximation can be sought via (i) sampling the model parameters with Markov Chain Monte Carlo (MCMC) inference [19] or (ii) variational inference (VI) method to find an equivalent distribution $q(\omega)$ by minimizing the divergence between this approximation and the true posterior $p(\omega)$ [16]. However, the number of weights and biases are large that the former approach (MCMC) will be computationally expensive in a DL neural network for air-pollutant forecast. Regarding to the latter method, Gal and Ghahramani have mathematically proved that the drop-out regularization during inference can approximate Bayesian inference in deep Gaussian processes [16]. As such, we apply the drop-out inference to formulate the proposed LSTM-BNN model for uncertainty quantification and accuracy enhancement for dust forecast.

Procedure 1 Forecast distribution inference

```
Input: Input sequences ($x_i \in X_{new}$),
Pretrained model $f(p(\omega|X_{train}))$
Initialize $Training_{Dropout} = True$
Initialize $Y_{forecast} = \emptyset$
Define forecast length: $Timesteps \in \mathbb{Z}$
Define distribution samples: $Samples \in \mathbb{Z}$
repeat
    for each $i \in Samples$ do
        $y_i^j = f(\omega_{i-Dropout}, x_i)$
    end for
    $y^j = Concatenate(y_i^j)$
until $j = Timesteps$
Output: $Y_{forecast} = Concatenate(y^j)$
```

Procedure 1 explains the distribution inference of model forecast for a number of future $Timesteps$ based on Bayesian approximation with the drop-out method. The key point of this method is to set the training status of drop-out layers to be active ($True$) to randomly vary the model's configuration (i.e., active nodes of each layer) and sample weights ($\omega_{i-Dropout}$) when making a prediction. The $Samples$ variable defines a number of iterative predictions y_i^j produced from the model $f(\omega_{i-Dropout})$ for an input sequence x_i. Then, all predictions are concatenated to form a distribution (y^j) of the forecast time step j^{th}. The model repetitively predicts until the final time step ($j = Timesteps$) and combines all the outputs to a sequence of the forecast distributions $Y_{forecast}$.

We practically find the forecast distributions of our data are converged to Gaussian distributions from at least 50 predictions. Hence, the value of $Samples$ is selected at 50 for all $Timesteps$. Then, the mean of Gaussian distributions can be inferred to a forecast value at each time step.

4 Results and discussion

This section presents the experiment results for the prediction of particulate matters $PM_{1.0}$, $PM_{2.5}$ and PM_{10} emitted on a construction site in Melrose Park

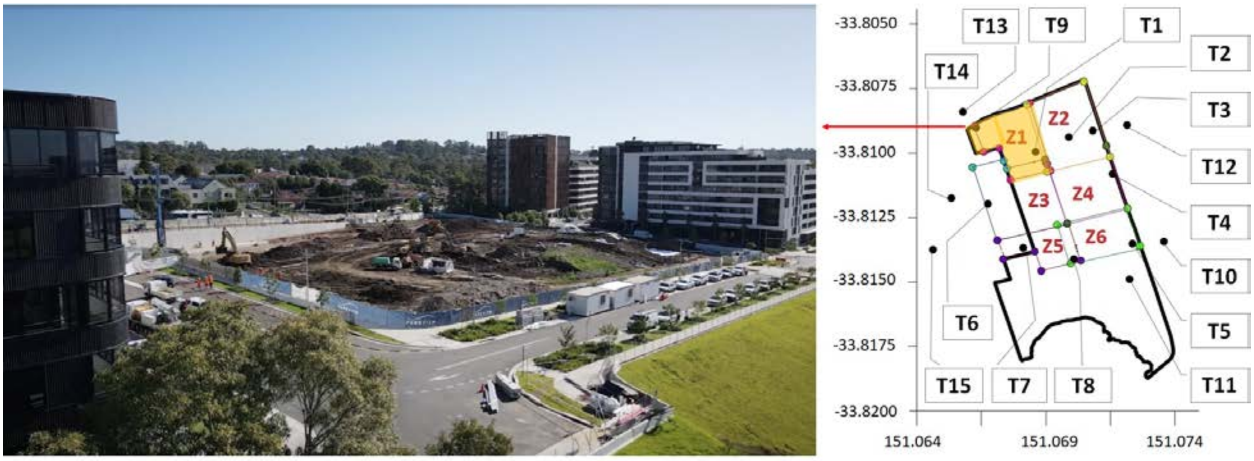

Figure 2. Construction site in Zone 1 (left) and the locations of 15 sensor motes for emission monitoring in Melrose Park and residential areas (right) [20, 3]

of the NSW suburb Paramatta North of coordinates ($33°49'11''S; 151°4'38''E$). The site is shown in Figure 2, where low-cost sensors are installed on the site as well as in residential areas surrounding it [3].

The forecast results are validated with ground-truth values in the early period of May 2020. Then, the performance of the proposed model (LSTM-BNN) will be compared with advanced DL models commonly used for time-series forecasts by popular statistical metrics.

4.1 Evaluation metrics

We use two error and two correlation metrics to evaluate the model's performance including:

- *The mean absolute error (MAE)*:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, \tag{15}$$

- *The root mean square error ($RMSE$)*:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{16}$$

- *The Pearson's correlation (r)*:

$$r = \frac{\sum(x_i - \hat{x}_i)(y_i - \hat{y}_i)}{\sqrt{\sum(x_i - \hat{x}_i)^2 \sum(y_i - \hat{y}_i)^2}}, \tag{17}$$

- *The coefficient of determination (R^2)*:

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}, \tag{18}$$

where y_i and $\hat{y}_i$ here are respectively the measured observations and forecast values of variable y at the i^{th} instant, (similarly to variable x), and n is the number of inspected samples. The lower values of $RMSE$ and MAE or higher values of r and R^2 indicate better performances.

4.2 Benchmark models

In order to validate the robustness of the proposed model, we benchmark LSTM-BNN with three popular models used to forecast time series:

- Gated recurrent unit (GRU) is considered as a lightweight version of LSTM with only two gates (update and reset gates). Both LSTM and GRU are robustly implemented for time series forecasting, but LSTMs are generally considered to be more powerful and better at capturing long-term dependencies in the data, while GRUs are considered to be faster and more computationally efficient [21].

- Bidirectional long short-term memory (BiLSTM) network is a variation of the LSTM network that processes the input sequence in two directions (i.e., forward and backward). The two LSTM layers are then concatenated and the output from both layers is combined to form the final output of the network [22].

- One-dimension convolution neural network (1D-CNN) model is recently implemented to forecast time series [23]. Although CNN models are popular in computer vision due to its robust capacity of feature extraction by 2D convolutional layers [17], special patterns of time series are also learned well with 1D convolutional filters. The deeper layers can learn more features of data and then can be reconstructed at the fully connected layer in the output of CNN models.

Both two RNN models (GRU and BiLSTM) share the same configuration with our proposed model (i.e., number of layers, nodes/units, drop-out proportion, learning rates, etc.) for a fair comparison. The CNN model is configured

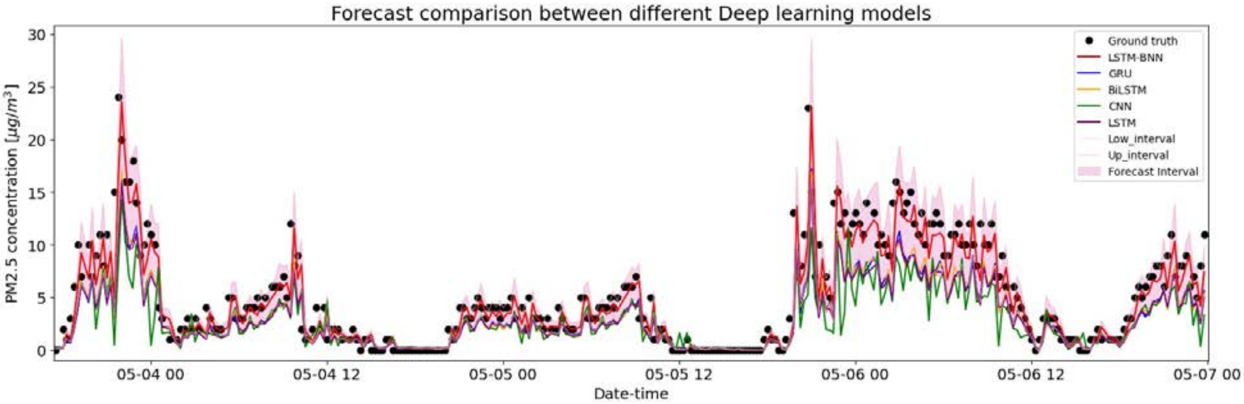

Figure 3. Forecast comparison between proposed model (LSTM-BNN) and benchmark models

by one 1D-CNN layer concatenated with a Flatten layer and two fully connected layers. Besides, the LSTM deterministic model will be accounted as the fourth benchmark model.

4.3 Experimental results

Table 1. Compared metrics of $PM_{1.0}$ forecasts

Models	**MAE**	**RMSE**	**Pearson_r**	R^2
GRU	1.573	2.894	0.929	0.287
BiLSTM	1.379	2.553	0.931	0.445
CNN	1.677	2.838	0.928	0.314
LSTM	1.508	2.839	0.927	0.314
LSTM-BNN	**0.658**	**1.320**	**0.929**	**0.852**

Table 2. Compared metrics of $PM_{2.5}$ forecasts

Models	**MAE**	**RMSE**	**Pearson_r**	R^2
GRU	1.211	2.335	0.891	0.703
BiLSTM	1.196	2.312	0.921	0.708
CNN	1.488	2.917	0.829	0.536
LSTM	1.229	2.378	0.915	0.692
LSTM-BNN	**0.886**	**1.731**	**0.946**	**0.837**

Table 3. Compared metrics of PM_{10} forecasts

Models	**MAE**	**RMSE**	**Pearson_r**	R^2
GRU	3.010	4.971	0.855	0.253
BiLSTM	3.065	5.077	0.849	0.220
CNN	3.138	4.988	0.835	0.244
LSTM	3.059	5.063	0.852	0.225
LSTM-BNN	**1.899**	**3.270**	**0.855**	**0.677**

Figure 3 depicts the comparison of $PM_{2.5}$ forecasts of all involved models with the ground truth over the period from 3^{rd} to 7^{th} of May 2020. In general, all models learn and follow well the patterns of fine particles. Presented in the red line, the means of forecast distributions from the LSTM-BNN model significantly fit to the ground truth (black dots) with the forecast interval (pink) covering most of measured values. GRU, BiLSTM and LSTM have similar forecast accuracies, while CNN model (green line) demonstrates strong fluctuations with the worst performance. There are underpredictions produced by all benchmark models at high concentrations. These results are probably caused by the volatile dynamics of particle concentrations, and the quality of sensory data is challenging the learning capacity of deterministic models. It indicates our hybrid LSTM-BNN model handles quite well aleatoric uncertainty. Besides, the quantity of training data in this case are quite limited to train DL models compared to other similar studies [7, 8]; hence, the robustness of LSTM-BNN reduces the epistemic uncertainty as lack of information.

Tables 1, 2 and 3 present the quantified error metrics of all models forecasting $PM_{1.0}$, $PM_{2.5}$ and PM_{10}, respectively. LSTM-BNN outperforms all benchmark models over all studied metrics. The significant improvements are recorded upto 56.3%, 27.9% and 37.9% in MAE for three particles compared to its deterministic model (LSTM). Although Pearson's r values are similar for all models, the coefficient of determination (R^2) of LSTM-BNN is nearly triple those calculated by all benchmark models for PM_{10} forecast (e.g., 0.677 as opposed to 0.225).

4.4 Estimation of missing information

Since missing data of LWSN operating in construction sites is inevitable [10], we conduct another experiment to verify the imputation capacity of the proposed model for recreating the continuous information. Here, a period of 02^{nd} - 05^{th} January 2020 in the Black Summer (Nov. 2019 - Jan. 2020) with a severe bushfire in Australia is selected for this experiment. The rationale for this choice is the significant levels of concentrations were measured in this period over the whole NSW state [12]. Hence, it would be impossible to evaluate the net emission from construction activities by using inferred values from the neighbor AQSs. Being confirmed by the Weather bureau of the Australian government, this was also the nation's hottest and driest summer, which disrupted the normal operation of some on-site sensors [3].

As shown in Figure 4, the full data (black line) are randomly removed some values with a defined proportion called the missing ratio to simulate the missing prob-

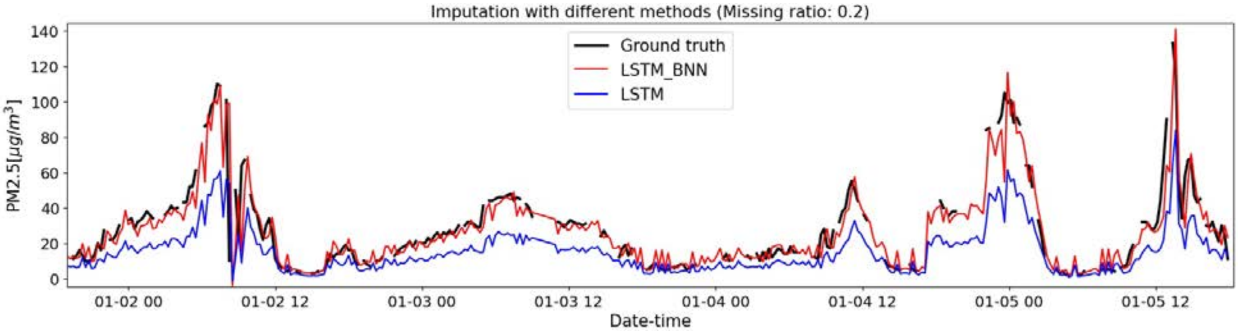

Figure 4. Imputations with LSTM-BNN and LSTM models for 20% randomly missing values of $PM_{2.5}$ over a period of 02^{nd} - 05^{th} January 2020

Table 4. Imputation capacity of LSTM-BNN compared with LSTM model with different missing ratios for $PM_{2.5}$

	MAE ($\mu g/m^3$)			R^2		
Missing ratios	**LSTM-BNN**	**LSTM**	**Improvement (%)**	**LSTM-BNN**	**LSTM**	**Improvement (%)**
0.1	6.261	12.723	50.8%	0.773	0.315	59.2%
0.2	6.521	12.627	48.4%	0.760	0.305	59.9%
0.3	8.324	13.175	36.8%	0.588	0.165	71.9%
0.4	9.032	13.353	32.4%	0.592	0.163	72.5%
0.5	9.668	13.544	28.6%	0.570	0.128	77.5%

lem. At a missing ratio of 0.2 (i.e., 20% of data being removed randomly), the imputation results for both probabilistic LSTM-BNN (red line) and deterministic LSTM (blue line) depict the trend of $PM_{2.5}$ being learned well by the remaining values of the LWSN and the data from AQSs. The LSTM-BNN model outperforms with smaller errors than the deterministic LSTM model over the whole studied period. We can see the estimated fluctuations of the LSTM-BNN model at the missing points due to less information from inputs, representing high epistemic uncertainty. When there are a higher number of absent values, more fluctuations are recorded. The performances of the two models at different missing ratios ranging from 0.1 to 0.5 are presented in Table 4. The statistical values in this table show huge improvements when imputing with the probabilistic LSTM-BNN model from 28.6% to 50.8% and from 59.2% to 77.5% for MAE and R^2, respectively.

5 Conclusion

We have presented a hybrid deep learning model for estimating dust emissions of construction sites by using LSTM and BNN with drop-out regularization during probabilistic inference. This model has merits of uncertainty quantification and mitigation for highly uncertain and limited data of LWSN, which addresses challenging issues of limited amount of on-site data and reliability of the data collected due to sensor noise [6]. The experimental results show significant improvements for important evaluation metrics in forecasting three types of particulate matters emitted during construction activities in a residential suburb. Since inferring the forecast values by the means of predictive Gaussian distributions requires sampling a high number of predictions in each future time step, the cost of computation will be more expensive than deterministic models. This will be our future work to develop a new inference method for reducing the number of predictive samples by estimating the practical distributions of model outputs. Besides, construction schedules are also one potential input to the proposed model that can be considered to further improve the prediction accuracy for on-site emissions concentrations.

References

[1] Huynh A.D. Nguyen, Lanh V. Nguyen, and Quang P. Ha. Iot-enabled dependable co-located low-cost sensing for construction site monitoring. In *Proc. 37th Int. Symposium on Automation and Robotics in Construction (ISARC)*, pages 616–624, Kitakyshu, Japan, 2020. doi:10.22260/ISARC2020/0086.

[2] Huynh A.D. Nguyen and Quang P. Ha. Robotic autonomous systems for earthmoving equipment operating in volatile conditions and teaming capacity: a survey. *Robotica*, pages 1–25, 2022. doi:10.1017/S0263574722000339.

[3] Huynh A.D. Nguyen and Quang P. Ha. Wireless sensor network dependable monitoring for urban air quality. *IEEE Access*, 10:40051–40062, 2022. doi:10.1109/ACCESS.2022.3166904.

[4] Denis Clement, Amir A Aliabadi, Jennifer Mackey, Jesse Thé, and Bahram Gharabaghi. Dust emissions management model for construction sites. *Journal of Construction Engineering and Management*, 147

(8):04021092, 2021. doi:10.1061/(ASCE)CO.1943-7862.0002121.

[5] NSW-DPE. Air quality guidance note - construction sites. On-line: www.environment.nsw.gov.au/resources/air/mod3p3construc07268.pdf, Accessed: 01/01/2023.

[6] Yayin Xu, Ying Zhou, Przemyslaw Sekula, and Lieyun Ding. Machine learning in construction: From shallow to deep learning. *Developments in the Built Environment*, 6:100045, 2021. ISSN 2666-1659. doi:10.1016/j.dibe.2021.100045.

[7] Taofeek D. Akinosho, Lukumon O. Oyedele, Muhammad Bilal, Ari Y. Barrera-Animas, Abdul-Quayyum Gbadamosi, and Oladimeji A. Olawale. A scalable deep learning system for monitoring and forecasting pollutant concentration levels on uk highways. *Ecological Informatics*, 69:101609, 2022. ISSN 1574-9541.

[8] Xiaohui Guo, Yuanfeng Wang, Shengqi Mei, Chengcheng Shi, Yinshan Liu, Lei Pan, Kai Li, Boqun Zhang, Junshan Wang, Zhiwu Zhong, and Minzhong Dong. Monitoring and modelling of pm2.5 concentration at subway station construction based on iot and lstm algorithm optimization. *Journal of Cleaner Production*, 360:132179, 2022. ISSN 0959-6526. doi:10.1016/j.jclepro.2022.132179.

[9] Moloud Abdar, Farhad Pourpanah, Sadiq Hussain, Dana Rezazadegan, Li Liu, Mohammad Ghavamzadeh, Paul Fieguth, Xiaochun Cao, Abbas Khosravi, U. Rajendra Acharya, V. Makarenkov, and S. Nahavandi. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*, 76:243–297, 2021.

[10] Santanu Metia, Huynh A. D. Nguyen, and Quang P. Ha. Iot-enabled wireless sensor networks for air pollution monitoring with extended fractional-order kalman filtering. *Sensors*, 21(16), 2021. ISSN 1424-8220. doi:10.3390/s21165313.

[11] Matthew Riley, John Kirkwood, Ningbo Jiang, Glenn Ross, and Yvonne Scorgie. Air quality monitoring in nsw: From long term trend monitoring to integrated urban services. *Air Quality and Climate Change*, 54(1):44–51, 2020. doi:10.3316/informit.078202598997117.

[12] Friends of the Earth Australia. Australia's black summer. On-line: www.foe.org.au/australia_black_summer, Accessed: 02/01/2023.

[13] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8): 1735–1780, 1997. doi:10.1162/neco.1997.9.8.1735.

[14] Abien Fred Agarap. Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*, 2018. doi:10.48550/arXiv.1803.08375.

[15] Stefan Wager, Sida Wang, and Percy S Liang. Dropout training as adaptive regularization. *Advances in neural information processing systems*, 26, 2013. doi:10.48550/arXiv.1307.1493.

[16] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *ICML*, pages 1050–1059. PMLR, 2016.

[17] Qiuchen Zhu, Hiep T. Dinh, Duong M. Phung, and Quang P. Ha. Hierarchical convolutional neural network with feature preservation and autotuned thresholding for crack detection. *IEEE Access*, 9:60201–60214, 2021. doi:10.1109/ACCESS.2021.3073921.

[18] Igor Kononenko. Bayesian neural networks. *Biological Cybernetics*, 61(5):361–370, 1989. doi:10.1007/BF00200801.

[19] Adam D Cobb, Atılım Günes Baydin, Ivan Kiskin, Andrew Markham, and Stephen J Roberts. Semi-separable hamiltonian monte carlo for inference in bayesian neural networks. In *Advances in Neural Information Processing Systems Workshop on Bayesian Deep Learning*, 2019.

[20] Melrose Park. Melrose park village construction update. On-line: https://melrosepark.com.au/news/melrose-park-village-construction-update, Accessed: 28/03/2023.

[21] Peter T Yamak, Li Yujian, and Pius K Gadosey. A comparison between arima, lstm, and gru for time series forecasting. In *Int. Conf. on Algorithms, Computing and Artificial Intelligence*, pages 49–55, 2019. doi:10.1145/3377713.3377722.

[22] Zhendong Zhang, Yongkang Zeng, and Ke Yan. A hybrid deep learning technology for pm 2.5 air quality forecasting. *Environmental Science and Pollution Research*, 28:39409–39422, 2021.

[23] Narendra Chaudhary, Sanchit Misra, Dhiraj Kalamkar, Alexander Heinecke, Evangelos Georganas, Barukh Ziv, Menachem Adelman, and Bharat Kaul. Efficient and generic 1d dilated convolution layer for deep learning. *arXiv preprint arXiv:2104.08002*, 2021.