Consensus Forecast of Rainfall Using Hybrid Climate Learning Model

Neethu Madhukumar, *Student Member, IEEE*, Eric Wang, *Member, IEEE*, Yi-Fan Zhang, *Member, IEEE*, and Wei Xiang, *Senior Member, IEEE*

***Abstract*—Rainfall event forecasting is prominently done using climate models (CMs) to produce multiple forecasts for the same rainfall event. The best forecast is complicated to find and hence has not yet been explored in the CMs. Recent advances in deep learning methods have provided an exceptional ability to investigate intricate weather patterns from big climate data. In this article, a hybrid climate learning model (HCLM) is proposed that utilizes both the CM and the deep learning models for improving the rainfall forecast. More specifically, a probabilistic multilayer perceptron (PMLP) network evaluates multiple forecasts from the CM-generated forecasts and selects the best one. The selected forecast is next passed onto a hybrid deep long short-term memory (HD-LSTM) network, which looks back and learns the relationship of the selected forecast with corresponding rainfall and temperature observations to produce the next-day rainfall forecast. The experimental results from various climate zones in Australia show that the HCLM outperforms existing state-of-the-art climate and deep learning models.**

***Index Terms*—Climate model (CM), hybrid model, long short-term memory, multilayer perceptron (MLP), rainfall forecast.**

I. Introduction

A RELIABLE model for predicting weather events can provide useful information for planning and management in many fields [1], [2]. A good weather prediction model should seamlessly link together diverse atmospheric changes for providing improved weather predictions. The majority of countries worldwide use climate models (CMs), such as the general circulation model (GCM), the Earth system models, and more to officially forecast weather patterns [3]–[5]. They are built upon differential equations based on the laws of physics, fluid motion, and chemistry, which are applied to the input data. On that account, forecasts can be considered as the numerical relationship between various atmospheric and environmental changes that result in the occurrence of climate events.

Manuscript received March 9, 2020; revised July 9, 2020 and September 30, 2020; accepted November 12, 2020. Date of publication November 26, 2020; date of current version April 23, 2021. This work was supported by the Beijing Natural Science Foundation under Grant L182032. *(Corresponding author: Wei Xiang.)*

Neethu Madhukumar and Eric Wang are with the College of Science and Engineering, James Cook University, Cairns, QLD 4878, Australia (e-mail: neethu.madhukumar@my.jcu.edu.au; eric.wang@jcu.edu.au).

Yi-Fan Zhang is with the Queensland Alliance for Agriculture and Food Innovation, University of Queensland, Brisbane, QLD 4072, Australia (e-mail: yifan.zhang@uq.edu.au).

Wei Xiang is with the School of Engineering and Mathematical Sciences, La Trobe University, Melbourne, VIC 3086, Australia (e-mail: w.xiang@latrobe.edu.au).

Digital Object Identifier 10.1109/JIOT.2020.3040736

The data to the CM are obtained from the Earth-observing satellites, the weather radars, and various climate sensors [6], [7]. These data then get processed using various scientific tools for climate impact, adaptation analysis, and weather forecasting thereby providing a higher representation of the input data. Over the years weather forecasting by the CM has greatly improved with the arrival of new technologies, integration of better theory, and increased computational power. Nevertheless, there is still a gap in finding the best forecast from multiple forecasts produced by the CM. The discovery of hidden weather patterns and relationships from the CM could significantly improve its performance. Deep learning techniques can capture underlying subtle functional relationships, unearth intricate structures from big climate data, and find the hidden patterns from complex CM forecast data. These features of climate and deep learning models make them ideal to form a teacher–student network for knowledge distillation from the CM to the deep learning model [8].

Neural networks have been used for predicting various weather attributes [9], [10]. Comparative analysis has indicated a higher performance accuracy for the neural network models than prominent data mining techniques for rainfall predictions [6]. A deep neural network (DNN) is a multilayer neural network with outstanding learning ability [11]. The success of deep learning models is due to its deep architectures. DNN-based approaches have become the most effective techniques for rainfall prediction due to its ability to progressively build higher-level representation from the weather data.

The DNN-based rainfall prediction approaches that utilize feedforward neural networks provide only a static network, that lack memory capacity [7], [12]. They do not consider temporal relationships. Comparative analysis has shown that a dynamic neural network with memory units exhibited improvement in predicting the rainfall than static feedforward neural networks [13], [14]. So from all the above works, it can be concluded that both the CM and the DNN-based approaches for rainfall predictions have been successful in forecasting certain rainfall features but fails to completely emulate the actual rainfall pattern. Henceforth, a new prediction method that utilizes both the CM and the DNN for the more accurate prognosis of rainfall events is proposed in this article.

The discovery of the best forecast pattern is difficult and consequently has not been investigated to improve the performance of CMs. Learning the best forecast pattern from the CM, and determining its relationship with other weather changes will improve the rainfall event prediction. We propose

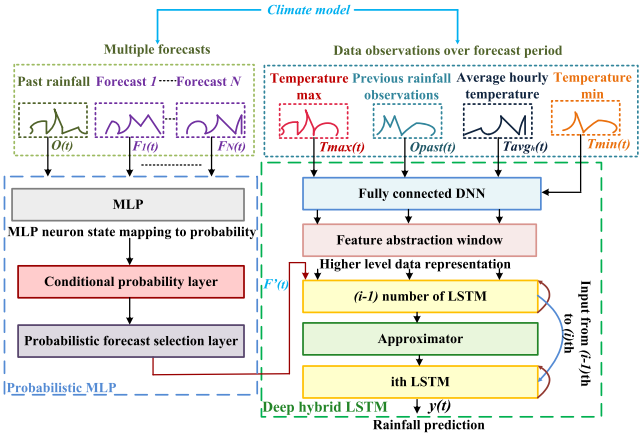


Fig. 1. Illustration of the HCLM structure. Multiple forecasts from the CM are input to the PMLP network, which conducts a conditional probability evaluation of the forecasts for finding the best of the *N* forecasts. The output of PMLP is input to HD-LSTM along with input data observations over the forecast period. HD-LSTM analyzes the PMLP forecast and the input data observed over the selected forecast period to predict rainfall.

a hybrid climate learning model (HCLM) that learns from the CM using deep learning. The HCLM shown in Fig. 1 is designed using a probabilistic multilayer perceptron (PMLP) and a hybrid deep long short-term memory (HD-LSTM) network for reprocessing the multiple forecasts produced by the CM for improved next-day rainfall prediction.

PMLP is used to rank and select the best forecast, and HD-LSTM finds the correlation between the PMLP-selected forecast and the corresponding data observations for predicting the next-day rainfall. The PMLP network evaluates the probability of each forecast and selects one forecast with the highest probability. The output from PMLP along with weather changes observed over the forecast period is analyzed by the HD-LSTM to produce the next-day rainfall forecast. The rainfall forecasts and the observation data from ten sites from different weather zones across Australia are selected to evaluate the performance of the HCLM. The main contributions of this article are summarized as follows.

1) A hybrid learning model that utilizes both the CM and the DNN is proposed. To the best of our knowledge, this is the first time that a hybrid DNN is used to refine multiple forecasts from the CM.
2) A PMLP network-based best forecast selection method is proposed, which analyzes, compares, and ranks multiple forecasts for different rainfall patterns.
3) A new improved HD-LSTM network is proposed for learning the relationship between the best forecasts from the CM and the corresponding data observations for producing the next-day rainfall forecast.
4) Experimental evaluation of the proposed HCLM model has shown that it outperforms existing state-of-the-art climate and deep learning models.

The remainder of this article is organized as follows. Section II lists the related work. Section III describes the proposed HCLM model. A list of all symbols used in this Section is given in Table I. Section IV gives the experimental results. Concluding remarks are drawn in Section V.

TABLE I
LIST OF SYMBOLS

Symbol	Meaning
t	$t \in \{1, 2, ..., T\}$, Time instant of each data from 1 to T.
n	$n \in \{1, 2, ..., N\}$, N Number of forecasts.
$F_n(t)$	Forecasts $n \in \{1, 2, ..., N\}$ at $t \in \{1, 2, ..., T\}$.
T_{sub}	Data blocks size.
W_{zv}^{m}	Weights from v^{th} neuron of layer $(m-1)$ to z^{th} neuron of layer m.
W	Collection of W_{zv}^{m} $\forall m, z, v$.
$P_j(t)$	Conditional Probability for $j \in \{1, 2, ..., N\}$, $\forall j \neq n$.
$F'(t)$	PMLP output.
U	Number of types of input observation data.
$x(t)$	Input data.
$h(t)$	Hidden vector sequence.
i	HD-LSTM current hidden state that receives the summary of past $(i-1)$ hidden states.
W_{xh}	Input to hidden weight matrix.
W_{hh}	Hidden to hidden weight matrix.
$b(t)$	Bias function.
z	Forecast lead time.
Wh, z	Weights from input data at time z to HD-LSTM deep input layer.
$W_{F'h}$	PMLP output to HD-LSTM hidden weight matrix.
ψ	Non-linear transition function between HD-LSTM layers.
k	$k \in \{1, 2, ..., K\}$ HD-LSTM layers.
$y(t)$	Final output.

II. RELATED WORK

A. Climate Models

Numerous types of CMs are used for weather forecasting. The energy balance CMs produce forecasts based on the relationship between the energy entering and leaving the atmosphere [3]. The radiative–convective CMs produce a forecast based on the changes in temperature and humidity in the atmosphere [15]. The GCM produces weather forecasts based on the flow of air in the atmosphere, water in the ocean, and transfer of heat in the atmosphere [4]. The Earth system models conduct an additional analysis of the changes in the biogeochemical cycle than the GCM [5]. Therefore, the rainfall forecasts can be considered as the numerical relationship with the changes in the atmosphere and environment. The forecast summary is usually provided by an expert using the complicated forecasting characteristics of numerical methods. This method has achieved acceptable results, but it often needs long-term experience and subjective judgment. Deep learning can learn and execute complex analytical tasks [16]. So they have high potential in summarizing the forecasting characteristics of the CM to further improve the rainfall prediction. Details of some existing weather predictions using deep learning models are discussed in the following section.

B. Deep Learning Models

Many research studies have been conducted on the application of deep learning models on weather predictions. An artificial neural network model with two hidden layers is used to predict rainfall [17]. This work indicated that the proposed neural network model performs better than the GCM. Similarly, another study used climate parameters and circulation factors from the CM as input to a 1-D deep convolutional neural network (1DD-CNN) [18]. This work illustrated the potential of deep learning models to objectively extract features to predict rainfall events from the CM. However, this study does not provide a comparison with any dynamic neural network models with memory units. This is critical since studies have indicated a higher performance for such networks over CNN in some types of time-series data [19]. A comparison with this will paint a clear picture of which network is more effective for rainfall.

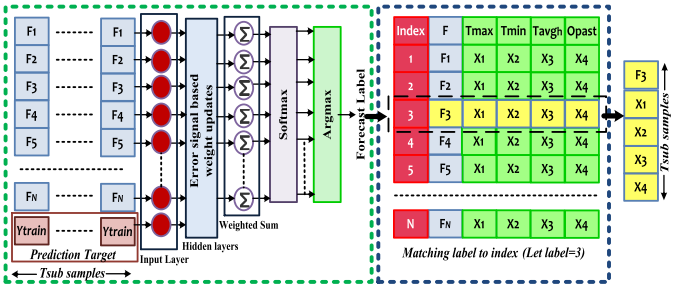


Fig. 2. Illustration of the PMLP structure. All forecasts and actual rainfall (red box) pass from the input layer to the hidden layer, which generates error signals for a given set of input forecasts. Errors are used by the nodes to update the values for each connection weight. A weighted sum of the inputs goes to the softmax layer, which provides a probabilistic estimate to the Argmax layer. The Argmax layer outputs the label of the best forecast, which is used to select the corresponding observation data for the best forecast.

Lin and Wu [20] proposed a network model using self-organizing-maps and multilayer perceptron (MLP), for predicting typhoon rainfall. Zhang *et al.* [21] proposed a deep belief network-based rainfall prediction. According to Mekanik *et al.* for spring rainfall prediction the most accurate performance was demonstrated by MLP [22]. Shi *et al.* [23] proposed a novel model based on LSTM to predict the structure and direction of clouds in the radar picture. It can better predict short-term rainfall than many popular deep learning models. In recent years, deep learning models have shown exceptional performance in weather predictions due to its hybrid deep architectures, such a work is done in [11], where a hybrid gated recurrent neural network is used for improved predictions for multiple meteorological events. Shi *et al.* [23] proposed a novel approach using hybrid LSTM for weather predictions by analyzing the structure and direction of clouds in the radar picture. Neural network-based climate data analysis has received a lot of attention in recent years to tackle several climate problems [2], [9], [10]. These works exhibit the effectiveness of deep hybrid neural networks for detection, classification, and predictions of weather events.

III. Hybrid Climate Learning Model

In HCLM, the PMLP learns the pattern of the best forecast from the CM over multiple data subsamples and selects the best forecast from the evaluation data. The PMLP output is fed to HD-LSTM as part of knowledge distillation. The use of the PMLP output label for data selection reduces the LSTM search space for sequential data analysis.

A. PMLP

In the PMLP structure illustrated in Fig. 2, N input forecasts are passed to the input layer of the PMLP. The N forecasts produced for the rainfall event by the CM are denoted by $F_n(t)$ for $n \in \{1, 2, \ldots, N\}$ at $t \in \{1, 2, \ldots, T\}$ and are passed on to the input layer of the PMLP for further refinement. $O_{\text{past}}(t)$ is the past rainfall observations. The pattern of each forecast changes over different time blocks, so the PMLP learns the relationship between $F_n(t)$ and $O_{\text{past}}(t)$ for multiple T_{sub} data blocks. Table II lists the monthly root-mean-square error (RMSE) of the 7-days forecasts for obtained from the Australian Bureau of Meteorology (BOM) for the Burdekin region in Queensland [24]. The data set was prepared by the BOM's evidence targeted automation team for the science to services program. In Table II, Forecast 1 is the first forecast produced seven days before the event, while Forecast 7 is the last forecast produced one day before the event. Forecast 7 is thought to be the best forecast as it has the shortest lead time to the actual rainfall event [5]. However, our analysis of the monthly forecast data (from 2015 to 2017 using [24]) shown in Table II indicated that this is not always true. For example, Forecast 7 predicted 6.4-mm rain and Forecast 1 predicted 0.9-mm rain for October 18, 2017, but the actual rain was 0.6 mm. So to forecast rain on October 18, 2017, information from Forecast 1 is more useful.

Table II illustrates that for each period there exists a forecast which is more reliable than the other forecasts. So the rest of the forecasts can be dismissed for that period. Hence, we propose to rank each forecast based on the patterns of each forecast and rainfall observations for multiple T_{sub} data blocks for selecting the best forecast. A probabilistic selection of any feature over different data blocks performs better than many benchmark methods [25], [26]. Therefore, such a selection is adopted. For the data set used in this article, $T_{\text{sub}} = 21$ days is chosen using the random search algorithm [28].

The PMLP analyzes all $F_n(t)$ for the rainfall patterns over multiple T_{sub} data blocks. After training, the PMLP for similar rainfall patterns and similar forecast inter-relationship selects one from $F_n(t)$ as its output, which it predicts to be the best forecast. Error signals for each hidden layer node are determined for the given inputs. The errors are used by the nodes to update the values for each connection weights. For the PMLP, let W_{lv}^{m} denote the values of the weights from the vth neuron of layer $(m-1)$ to the lth neuron of layer m, and W be the collection of W_{lv}^{m} $\forall l, v, m$. So for the given set of inputs, the outputs can be represented as $S_n(F_n(t); W)$, a function of inputs and weights. The PMLP softmax layer estimates the conditional probability of all forecasts. Using [25], this can be expressed as

$$P_j(t) = P(F_j(t)|F_n(t); W) \text{ for } j \in \{1, 2, \ldots, N\} \ \forall j \neq n = \frac{e^{S_j(F_n(t);W)}}{e^{S_1(F_n(t);W)} + e^{S_2(F_n(t);W)} + \cdots + e^{S_N(F_n(t);W)}} \quad (1)$$

where $P(F_j(t)|F_n(t); W)$ is the conditional probability of forecast j for the given set of inputs $F_n(t)$ and the given set of weights W. The PMLP ranks and selects the best forecast using (1) over T_{sub} blocks. Let $F'(t)$ be the PMLP selected forecast, then

$$F'(t) = F_n(t), \text{ if } P_n(t) > P_j(t) \quad (2)$$

where $j \in \{1, 2, \ldots, N\}$ $\forall$ $j \neq n$. Using (2), the best forecast $F'(t)$ is selected from $F_n(t)$. The PMLP output $F'(t)$ is given as input to the HD-LSTM for temporal analysis with the temperature and the rainfall observations during the forecast time. For the given rainfall events, the temporal changes in observation data at the forecast period $t - z$, where z is

TABLE II
MONTHLY RMSE OF SEVEN DAYS FORECASTS OF BOM AUSTRALIA IN 2017. FORECAST i PRODUCED $(8-i)$ DAYS BEFORE THE EVENT

Month	January	February	March	April	May	June	July	August	September	October	November	December
Forecast 1	10.485	27.955	12.086	13.916	1.575	1.040	0.862	6.334	**0.398**	3.171	2.313	**2.316**
Forecast 2	9.886	26.601	12.149	14.358	1.580	1.662	0.659	2.460	0.560	3.558	3.602	2.448
Forecast 3	8.693	25.544	11.174	**11.252**	1.554	**0.604**	0.509	2.239	1.093	4.039	**2.132**	2.437
Forecast 4	8.752	25.026	9.097	14.913	**1.391**	0.976	0.705	2.400	0.916	3.312	2.493	2.375
Forecast 5	8.369	24.992	8.710	15.864	1.563	0.706	0.723	**0.514**	0.456	**3.128**	2.753	2.454
Forecast 6	7.184	**24.266**	7.921	12.776	1.500	1.071	0.691	2.198	0.772	4.872	2.410	2.431
Forecast 7	**6.887**	25.388	**7.796**	11.368	1.855	0.610	**0.420**	2.236	1.100	4.522	2.500	2.381

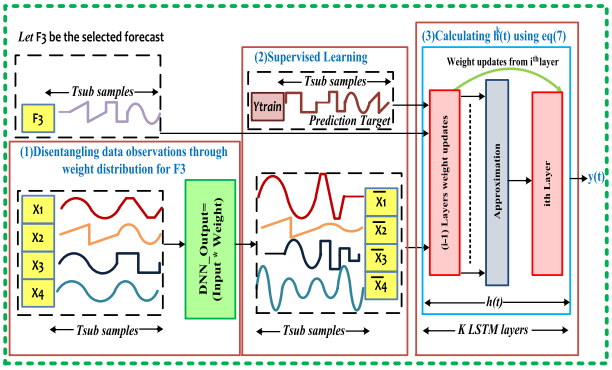


Fig. 3. Illustration of the HD-LSTM structure. The learning process is implemented in three steps: 1) DNN disentangles data observations through weight distribution; 2) supervised learning uses PMLP forecast and observation data with real rainfall as the prediction target; and 3) input state transition by the network through $h^k(t)$ from (7), and the network predicts the output using (8).

the forecast lead time, affect its decisions. The PMLP argmax layer provides the label of the best forecast for matching with the index of the data set to select the observation data. The HD-LSTM analyzes the relationship of the selected forecast with the observations. The HD-LSTM is detailed as follows.

B. HD-LSTM

In the HD-LSTM structure illustrated in Fig. 3, the output of PMLP is given to the HD-LSTM for further analysis. The HD-LSTM learns the relationship between the PMLP output and the observation data to predict the next-day rainfall. For the CM used, the data observations are the maximum temperature $T_{\max}(t)$, the minimum temperature $T_{\min}(t)$, the past rainfall $O_{\text{past}}(t)$, and the average hourly temperature $T_{\text{avg}_h}(t)$. For the ease of derivation, we denote $x(t)$ as the input data for the HD-LSTM and U types of data are denoted by $x_u(t)$. Here for derivation, we use $[T_{\max}(t), T_{\min}(t), O_{\text{past}}(t), T_{\text{avg}_h}(t)] = [x_1(t), x_2(t), x_3(t), x_4(t)]$. For simplifying derivation, we first derive for one type of observation, denoted by $x(t)$.

For the selected forecast $F'(t)$ by the PMLP, the HD-LSTM applies a fully connected DNN at its input to select the data observations over the best forecast period. This data is passed on to the K LSTM layers. The network then analyzes the temporal correlation between the forecast and changes in the observation data. To make the comparison further easier for the HD-LSTM hidden layers, a summary of the changes in the $(i-1)$ hidden sates are provided to the ith hidden state. This implementation aids the network to look back at the summary of historical changes with current changes. The HD-LSTM within the HCLM progressively builds up a higher-level representation of input weather data. From [27], the equation of the LSTM can be expressed as

$$h(t) = H(W_{xh}x(t) + W_{hh}h(t-1) + b_h(t)) \tag{3}$$

where the input sequence is $x(t)$, the hidden vector is $h(t)$, W_{xh} denotes the input–hidden weight matrix, W_{hh} denotes the hidden to hidden weight matrix, $b_h(t)$ is the hidden bias function, and H is the hidden layer function.

A fully connected DNN at the input of the HD-LSTM implements a selection window. It selects the data observations x, which are the same as data observations over the best forecast period of z. The DNN matches the selected forecast label to index of observation data set. This assists the network to establish relationships between the temporal changes in the PMLP forecast and observations. This is highly necessary for establishing the puzzling forecast relationships with other climatic changes $x(t)$ over forecast period z. The DNN disentangling data observations through weight distribution for the selected forecast. So for the proposed network, $W_{xh}x(t)$ in (3) becomes $W_{\bar{x}h}\overline{x(t)}$

$$W_{\bar{x}h}\overline{x(t)} = \sum_{z=a}^{b} W_{h,z}x(t-z) \tag{4}$$

where $a \le z \le b$ defines the size of the selection window that ranges between 0 to 7, t corresponds to the time of the actual event, and z corresponds to the forecast time of $F'(t)$. It is the window that the network looks back to learn the temporal correlation with $F'(t)$. $W_{h,z}$ is the input weight from the deep input transcription layer over the range z. $W_{\bar{x}h}\overline{x(t)}$ and $F'(t)$ are analyzed by the HD-LSTM to learn what changes in $x(t)$ over forecast interval led to the best forecast prediction.

For learning the relationship of the current dynamic changes with the past CM data, an approximator between $(i-1)$ hidden states and the ith hidden state is implemented in the HD-LSTM. This provides a useful summary of the past to the present hidden states along with the current input. Thus, enabling the HD-LSTM to establish the relationship between the two. This is implemented using a vanilla neural network. Hence for the HD-LSTM used in this article, $h(t)$ in (3) can be expressed as

$$\begin{aligned} h(t) = \psi_h[W_i\psi_{i-1}[\cdots[\psi_1[W_{hh}h(t-1) + W_{\bar{x}h}\overline{x(t)} \\ + W_{F'h}F'(t)]]]] + b_h(t) \end{aligned} \tag{5}$$

where $F'(t)$ is obtained from the PMLP, $W_{F'h}$ denotes the PMLP output to the HD-LSTM hidden weight matrix, $W_{\bar{x}h}\overline{x(t)}$

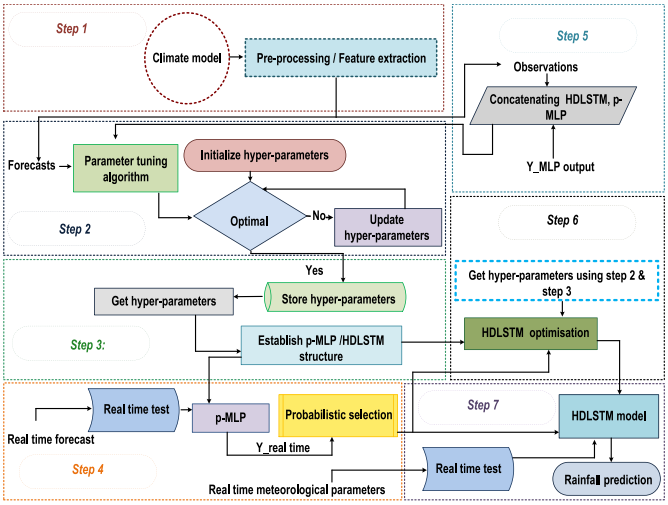


Fig. 4. Process of establishing HCLM is done in seven steps: 1) preprocessing and feature extraction; 2) PMLP structure selection and parameter tuning; 3) PMLP model optimization; 4) forecast selection using PMLP; 5) the PMLP is concatenated with the HD-LSTM for temporal analysis of forecasts with observations over forecast interval; 6) HD-LSTM model optimization is done using steps 2 and 3; and 7) rainfall prediction is obtained.

is obtained using (1), ψ_h is the nonlinear function obtained using the vanilla neural network, and W_i is the weight to the vanilla neural network between $(i-1)$ hidden states and the ith hidden state. For $k \in \{1, 2, \ldots, K\}$ layers, the hidden vector sequence $h(t)$ in (5) becomes $h^k(t)$. This can be expressed as

$$h^k(t) = \psi_h[W_i\psi_{i-1}[\cdots[\psi_1[W_{h^kh^k}h^k(t-1) + W_{\overline{x}h^k}\overline{x(t)} + W_{F'h^k}F'(t)]]]] + b_h^k(t). \quad (6)$$

Substituting for u types of data observations to (6), then

$$h^k(t) = \psi_h[W_i\psi_{i-1}[\cdots[\psi_1[W_{h^kh^k}h^k(t-1) + (W_{\overline{x_1}h}\overline{x_1(t)} + \cdots + W_{\overline{x_u}h}\overline{x_u(t)}) + W_{F'h^k}F'(t)]]]] + b_h^k(t). \quad (7)$$

The network computes all the K hidden states. Once the value of $h^k(t)$ at $k = K$ is obtained, the network output $y(t)$ can be calculated as

$$y(t) = W_{h^Ny}h(t)^N + b_y(t) \quad (8)$$

where $b_y(t)$ is the output bias function. The process of building this structure is discussed in the following section.

C. HCLM Implementation

The HCLM was developed using Python 3.5 with TensorFlow as the backhand. A detailed explanation of the main steps illustrated in Fig. 4 for developing the HCLM is given as follows.

1) Data Preprocessing and Feature Extraction: The PMLP inputs are the multiple rainfall forecasts and past rainfall observations. They are inputted to the PMLP for selecting the best forecast $F'(t)$ from $F_n(t)$. The training set for the PMLP can be denoted as $Q = \{F_n(t), t = 1, 2, \ldots, T\}$ for T number of training data. The PMLP output $F'(t)$ and $x_u(t)$, the U types of observation data are passed to the HD-LSTM. The training set for the HD-LSTM can be denoted as $R = \{(x_u(t), F'(t)), t = 1, \ldots, T\}$ for T number of data.

2) Structure Selection and Parameter Tuning: Parameter tuning is tedious and costly as the network goes through numerous iterations according to the performance for different sets of hyperparameters. The random search method exhibits the same performance as many common hyperparameter optimization methods with lower computation cost [28]. Hence, we have chosen the random search method. The key hyperparameters optimized are the length of the input sequence and the number of nodes in the hidden layer. A dropout of 20% is used to prevent overfitting.

3) PMLP Model Optimization: Once tuning is completed, optimal parameters are obtained. Next, the model undergoes training to find the best forecasts for the given inputs. PMLP learns through the backpropagation of errors by searching the data set containing all forecasts to find the best of forecasts that enhance the performance. This provides the network with efficient learning updates with high-quality gradients.

4) Real-Time Prediction: Once training is completed the PMLP will predict $F'(t)$; the best forecast by analyzing the given set of multiple forecasts $F_n(t)$. The PMLP uses (2) for selecting the best forecast as its output.

5) Model Concatenation: $F'(t)$ is the best forecast obtained using the PMLP. For U types of weather data changes observed over the forecast period, the observations $x_u(t)$ undergo temporal analysis with the best of forecasts $F'(t)$. For this purpose, the PMLP output is concatenated to the HD-LSTM.

6) HD-LSTM Model Optimization: Steps two and three are applied to the HD-LSTM for model optimization. For rainfall events generally, it is thought that recent weather data is more important for prediction. However, older data can help the model in recognizing general trends and movements. So the selection of sequence length is done for the rainfall data.

7) Rainfall Event Prediction: For the given data observations $x_u(t)$ and $F'(t)$, the best forecast selected by the PMLP, the HD-LSTM predicts the rainfall for the next day.

Fig. 5 illustrates how the PMLP selects the best forecast, and HD-LSTM correlates with the observations after completing the above seven steps of establishing the HCLM. The difference between the seven forecasts and the weighted sum of the past 21 days' PMLP inputs is calculated. The one with the least error is selected as the best forecast. For the selected forecast, the corresponding $T_{\max}(t)$, $T_{\min}(t)$, and $T_{avgh}(t)$ values are taken. From the list of 21 $T_{\max}(t)$, $T_{\min}(t)$, and $T_{avgh}(t)$ values, those that do not match with the output time observations selected by the PMLP are removed by the DNN. The $O_{\text{past}}(t)$ values corresponding to the DNN retained $T_{\max}(t)$, $T_{\min}(t)$, and $T_{avgh}(t)$ are taken for further examination. Next, the remaining values of $T_{\max}(t)$, $T_{\min}(t)$, $T_{avgh}(t)$, and $O_{\text{past}}(t)$ are passed on to the LSTM for analysis. For the given sequence, the LSTM looks for a row with the minimum difference between $O_{\text{past}}(t)$ and the PMLP selected rainfall forecast, which is chosen as the output. The values of the selected row are weighted to scale the data to predict the rainfall for the target day. Hence to summarize, the HCLM model uses a PMLP network to select the best forecast. The best forecast produced by the PMLP is compared with the other observed weather changes over the forecast period by the HD-LSTM to predict the next-day rainfall. The HCLM is tested in Section IV.

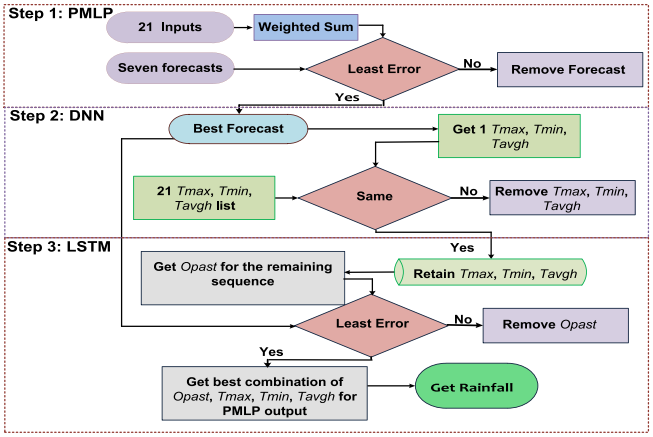


Fig. 5. Illustration of the HCLM working mechanism. 1) The forecast having the least error with the weighted sum of the past 21 days is chosen as the best forecast by PMLP; 2) DNN removes all $T_{\max}(t)$, $T_{\min}(t)$, and $T_{avgh}(t)$ data from the list of 21 sequences that does not match with the observations corresponding to PMLP selection and retains $O_{\text{past}}(t)$ for the remaining values; and 3) HD-LSTM analyze the DNN input sequence and selects the one with minimal error with the PMLP output. For the final selected combination of $T_{\max}(t)$, $T_{\min}(t)$, $T_{avgh}(t)$, $O_{\text{past}}(t)$, and PMLP best forecast, the LSTM predicts the rainfall.

IV. Experimental Results

To evaluate the HCLM, we test it on the data set from the BOM of Australia [24]. According to the BOM, there exist six major climate zones, namely, the equatorial, the tropical, the subtropical, the desert, the grassland, and the temperate zones. We compare our model with the best of BOM forecasts produced using the CM [5], the LSTM which is the most widely used neural network for time-series data analysis [14], [29] and two recent related works 1DD-CNN [18] and DBN-GRU [11]. For real data analysis, ten stations that span across six climate zones are considered for model training and preliminary verification by all the models (123 640 data points). For testing, we used rainfall data from June 2017 to May 2018. For training the HCLM, LSTM, 1DD-CNN, and DBN-GRU, we used data from May 2015 to May 2017. We used seven forecasts for the daily rainfall and the observation data, i.e., $O_{\text{past}}(t)$, $T_{\max}(t)$, $T_{\min}(t)$, and $T_{avg_h}(t)$. HCLM, LSTM, 1DD-CNN, and DBN-GRU use this data to predict the next-day rainfall. Daily rainfall forecasts from the CM are produced over a 24-h period back to back, where the last forecast is available only one day before the event. Therefore, this study aims at only the next-day prediction.

The 1DD-CNN [18] is used for comparison as it uses observational data from the BOM as input for rainfall prediction. The simulation results in [18] indicate a higher performance than BOM's forecasts. Therefore, this provides a fair comparison to our approach. While [11] uses a hybrid gated recurrent unit (DBN-GRU) model to analyze time-series meteorological data. This comparison is done since LSTM and GRU introduce additional gating components that handle vanishing gradients problem in traditional RNN for time-series data analysis. A comparison will help understand which of the two can better analyze time-series weather data. Hence, we find these works very important for comparison.

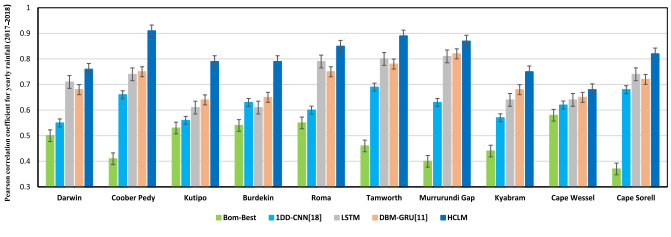


Fig. 6. Pearson correlation values between the BOM-Best, LSTM, 1DD-CNN, DBM-GRU, and HCLM predicted rainfalls. The HCLM, LSTM, 1DD-CNN [18], DBM-GRU [11], and BOM-Best exhibit average correlation of 0.81, 0.72, 0.61, 0.70, and 0.48, respectively.

TABLE III
Rainfall Prediction Error Comparison

SE	0.67	0.58	0.64	0.52	0.47
Location	**BOM**	**LSTM**	**1DD-CNN[18]**	**DBN-GRU[11]**	**HCLM**
Darwin	8.64	7.56	8.20	6.89	6.15
Coober Pedy	1.42	1.26	1.30	0.96	0.85
Kutipo	3.86	2.40	2.82	2.22	1.79
Burdekin	5.09	4.92	5.28	4.28	3.64
Roma	3.93	2.94	3.21	2.55	1.80
Tamworth	3.55	2.73	2.96	2.62	2.20
Murrurundi Gap	3.49	2.57	2.94	2.50	2.09
Kyabram	3.00	2.25	2.39	2.15	1.56
Cape Wessel	6.48	4.83	5.51	4.46	3.41
Cape Sorell	6.77	4.00	5.05	3.62	2.63
Average	4.62	3.55	4.00	3.23	**2.61**

Note: SE stands for standard error. The data reported in this table are for the 10 stations during 2017-2018.

A. Statistical Analysis

The accuracy of daily weather prediction for 12 months is compared using the Pearson correlation coefficient. The closer the value is to the $+1$, the better the forecasts. The bar graph in Fig. 6 illustrates the Pearson correlation coefficient values obtained for ten stations over one year. The graph illustrates a higher temporal correlation of the HCLM than [11], [18], the LSTM, and the BOM. There is a significant improvement in the fitting of time-series fluctuations of the HCLM predicted rainfall. For the average temporal correlation for ten stations for one year calculated using the Pearson correlation, the HCLM exhibits an average correlation of 0.81, LSTM exhibits an average correlation of 0.72, [18] exhibits an average correlation of 0.61, [11] exhibits an average correlation of 0.70, and the BOM exhibits an average correlation of 0.48.

In weather predictions, forecasting extreme rainfall events is a major challenge. Fig. 7 shows the extreme rainfall events for four locations located in high rainfall zones between June 2017 to May 2018. The comparison results show that the HCLM can outperform the BOM, the LSTM, [18], and [11]. Quantitative measurements are considered for synthesizing the effectiveness of the performances over 12 months for ten stations. Here, RMSE is used to calculate the accuracy of the HCLM. The closer the RMSE values are to zero, the better the forecasts. Fig. 8 shows the RMSE for rainfall predictions for ten locations. The RMSE curves show that the proposed HCLM has a lesser error in predicting the rainfall amount in comparison to the BOM and the deep learning models. Table III gives the quantitative RMSE values for the ten stations. It can be seen from this table that the average RMSE of the HCLM is the

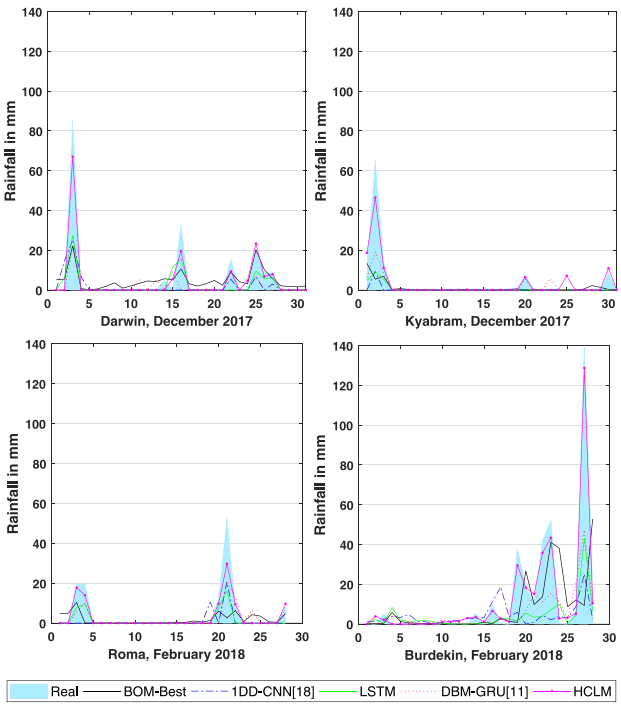


Fig. 7. Rainfall forecast results of four stations that experienced the highest rainfall events between June 2017 to May 2018. Graphically compared HCLM with LSTM, 1DD-CNN [18], DBN-GRU [11], and BOM-Best.

TABLE IV
COMPARISON OF PROBABILITY OF RAINFALL OCCURRENCE

SE	0.029	0.020	0.021	0.018	0.021
Location	**BOM**	**LSTM**	**1DD-CNN[18]**	**DBN-GRU[11]**	**HCLM**
Darwin	71.51%	74.25%	80.00%	78.63%	**90.68%**
Coober Pedy	86.84%	81.91%	89.31%	83.54%	**97.26%**
Kutipo	80.68%	84.18%	80.79%	85.93%	**90.57%**
Burdekin	78.15%	83.95%	79.34%	81.29%	**89.23%**
Roma	85.20%	91.23%	85.21%	90.14%	**92.33%**
Tamworth	69.31%	80.34%	77.15%	82.33%	**84.65%**
Murrurundi Gap	70.68%	81.23%	79.82%	80.95%	**86.57%**
Kyabram	68.76%	79.29%	76.87%	79.67%	**83.81%**
Cape Wessel	66.30%	77.83%	74.56%	78.31%	**80.43%**
Cape Sorell	56.16%	66.98%	64.10%	68.01%	**73.92%**
Average	73.36%	80.12%	78.72%	80.88%	**87.04%**

Note: SE stands for standard error. The data reported in this table are for the 10 stations during 2017-2018.

smallest among all approaches, indicating that the HCLM has higher rainfall prediction accuracy.

The rainy and nonrainy days occurrence prediction accuracy test was carried out for the ten locations. Table IV shows the prediction results. The results show that there is a notable improvement in the prediction accuracy of the HCLM in comparison with the BOM, the LSTM, the 1DD-CNN [18], and DBN-GRU [11]. There is an improvement of 18.65% for the HCLM over the best of BOM seven days forecasts. Also the HCLM is 8.64% more accurate than the LSTM. With the 1DD-CNN [18] there is an improvement of 10.57%. And to the DBN-GRU model in [11], there is an improvement of 7.10%. Hence from the results, it is apparent that the HCLM model is highly accurate in predicting the chance of occurrence of rainy days in comparison to the BOM, LSTM, [18], and [11].

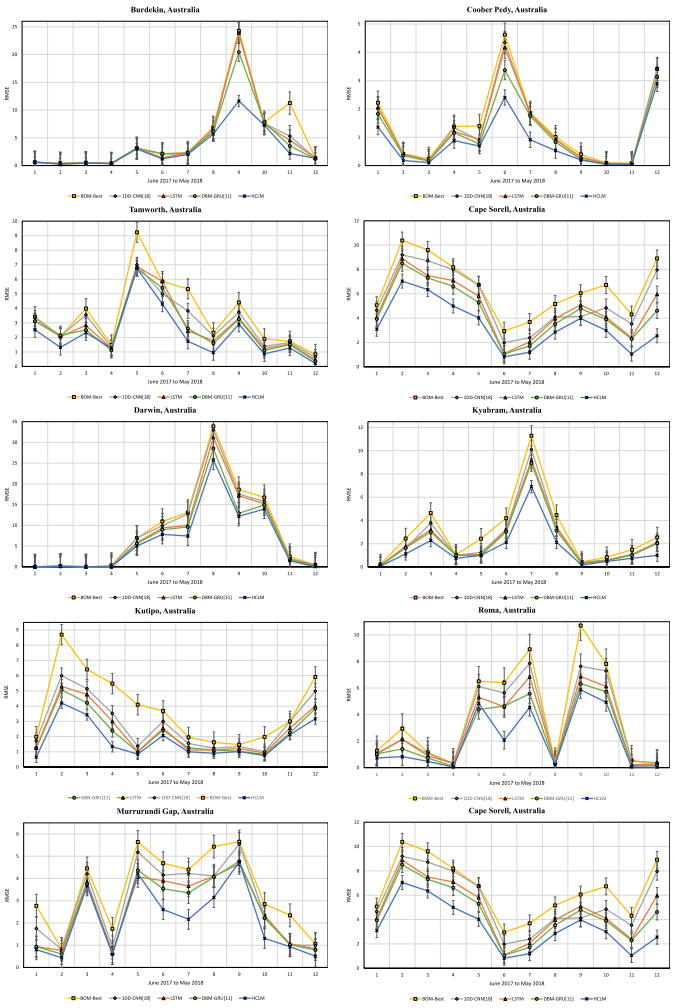


Fig. 8. Comparison of the RMSE for ten locations across all major climate zones in Australia for 12 months (source: BOM).

B. Ablation Studies

To verify the effectiveness of the individual network modules of the HCLM, ablation studies were conducted using seven forecasting approaches: 1) *PMLP:* the performance of the PMLP module with the seven forecasts as input; 2) *HD-LSTM+(All):* HD-LSTM with all the forecasts and observations as input; 3) *HD-LSTM+(Forecast* $7+$ *Obs):* HD-LSTM with the last forecast and all observations as input; 4) *HD-LSTM-*(O_{past}*):* HD-LSTM without $O_{\text{past}}(t)$ as input; 5) *HD-LSTM-*($T_{\max}$*):* HD-LSTM without $T_{\max}(t)$ as input; 6) *HD-LSTM-*($T_{\min}$*):* HD-LSTM without $T_{\min}(t)$ as input and 7) *HD-LSTM-*(T_{avg_h}*):* HD-LSTM without $T_{avg_h}(t)$ as input. Approaches 4–7 use all the BOM predictions. Table V shows the ablation study comparison with the HCLM.

The higher RMSE of PMLP than HCLM suggests that PMLP alone does not achieve the same performance as that of HCLM. Approach 2 shows that the HD-LSTM module has a smaller RMSE than PMLP, signifying that linking changes in observation data with the CM forecasts is essential. However, the higher RMSE of Approach 2 than HCLM suggests that rather than more training data, better training data will improve

TABLE V
COMPARISON OF ABLATION STUDIES

Forecasting Approach	HCLM	PMLP	HD-LSTM+(All)	HD-LSTM+(Forecast 7+Obs)	HD-LSTM-(O_{Past})	HD-LSTM-($T_{\max}$)	HD-LSTM-($T_{\min}$)	HD-LSTM-($T_{\mathrm{avg_h}}$)
RMSE	**2.61**	4.97	3.49	7.93	6.38	6.17	5.82	3.51
SE	**0.47**	0.43	0.62	0.89	0.71	0.76	0.63	0.48

network learning. So the use of PMLP to select the best forecast may have boosted the performance of the HCLM. The higher RMSE of approach 3 denotes that the final forecast alone cannot improve the performance of HD-LSTM.

Approaches 4–7 remove the observation data one by one from HD-LSTM, so as to identify the removal of which observations has the most effect. Table V shows that the removal of $O_{\mathrm{past}}(t)$ has the most effect on HD-LSTM, probably because there was a learning gap with the HD-LSTM unable to relate the past rainfall trends with the forecast. The removal of $T_{avg_h}(t)$ seems to have the least effect. This could be because the model was able to learn the average hourly temperature trends from $T_{\max}(t)$ and $T_{\min}(t)$ data. Also, approaches 4–7 do not outperform HCLM, further indicating that directly inputting all forecasts does not enhance the model learning. This must be the case since the forecast errors add up and misguide the model. So these results demonstrate that the PMLP and HD-LSTM work collaboratively to enhance the performance of the HCLM.

V. CONCLUSION

In this article, an HCLM is proposed. The HCLM builds its future rainfall predictions on the multiple forecasts produced by the CM for the rainfall events. The HCLM is designed using a PMLP network and an HD-LSTM network. The PMLP selects the best forecast from the multiple CM forecasts and the HD-LSTM finds the relationship of the selected forecast with the temperature and the rainfall observations during the forecast time. The evaluation was conducted on ten different locations that span across six major climate zones in Australia. Results showed that the HCLM predicted rainfall is more similar to the real rainfall than the CM and deep learning models tested, for average yearly results in all the locations analyzed. The potentiality of the model predicted rainfall has been established in terms of ablation studies, RMSE, and Pearson correlation. Hence, it can be concluded that there is a significant improvement by the HCLM to predict the rainfall over the popular climate and deep learning models.

REFERENCES

[1] N. Tang, S. Mao, Y. Wang, and R. M. Nelms, "Solar power generation forecasting with a LASSO-based approach," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 1090–1099, Apr. 2018.

[2] A. Koesdwiady, R. Soua, and F. Karray, "Improving traffic flow prediction with weather information in connected cars: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 9508–9517, Dec. 2016.

[3] A. H. Sodergren, A. J. Mcdonald, and G. E. Bodeker, "An energy balance model exploration of the impacts of interactions between surface albedo, cloud cover and water vapor on polar amplification," *Climate Dyn.*, vol. 51, pp. 1639–1658, Sep. 2018.

[4] E. Price, J. Mielikainen, M. Huang, B. Huang, H. A. Huang, and T. Lee, "GPU-accelerated long-wave radiation scheme of the rapid radiative transfer model for general circulation models (RRTMG)," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 8, pp. 3660–3667, Aug. 2014.

[5] D. Hudson *et al.*, "ACCESS-S1: The new bureau of meteorology multi-week to seasonal prediction system," *J. Southern Hemisphere Earth Syst. Sci.*, vol. 67, no. 3, pp. 132–159, Dec. 2017.

[6] R. S. Sangari and M. Balamurugan, "A survey on rainfall prediction using data mining," *Int. J. Comput. Sci. Mobile Appl.*, vol. 2, no. 2, pp. 84–88, Feb. 2014.

[7] G. Xu and V. Chandrasekar, "Operational feasibility of neural-network-based radar rainfall estimation," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 1, pp. 13–17, Jan. 2005.

[8] Y. Tu, J. Du, and C. Lee, "Speech enhancement based on teacher–student deep learning using improved speech presence probability for noise-robust speech recognition," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 27, no. 12, pp. 2080–2091, Dec. 2019.

[9] W. Wu and M. Peng, "A data mining approach combining *K*-means clustering with bagging neural network for short-term wind power forecasting," *IEEE Internet Things J.*, vol. 4, no. 4, pp. 979–986, Aug. 2017.

[10] M. Chen, G. Zeng, K. Lu, and J. Weng, "A two-layer nonlinear combination method for short-term wind speed prediction based on ELM, ENN, and LSTM," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6997–7010, Aug. 2019.

[11] Y. Cheng, X. Zhou, S. Wan, and K. R. Choo, "Deep belief network for meteorological time series prediction in the Internet of Things," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4369–4376, Jun. 2019.

[12] F. Beritelli, G. Capizzi, G. Lo Sciuto, C. Napoli, and F. Scaglione, "Rainfall estimation based on the intensity of the received signal in a LTE/4G mobile terminal by using a probabilistic neural network," *IEEE Access*, vol. 6, pp. 30865–30873, 2018.

[13] K. C. Luk, J. E. Ball, and A. Sharma, "An application of artificial neural networks for rainfall forecasting," *Math. Comput. Model.*, vol. 33, pp. 683–693, Mar./Apr. 2001.

[14] Y. Yu, J. Cao, and J. Zhu, "An LSTM short-term solar irradiance forecasting under complicated weather conditions," *IEEE Access*, vol. 7, pp. 145651–145666, 2019.

[15] W. C. Wang and P. H. Stone "Effect of ice-albedo feedback on global sensitivity in a one-dimensional radiative-convective climate model," *J. Atmosp. Sci.* vol. 37, no. 3, pp. 545–552, Mar. 1980.

[16] R. Ranjan *et al.*, "Deep learning for understanding faces: Machines May be just as good, or better, than humans," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 66–83, Jan. 2018.

[17] J. Abbot and J. Marohasy, "Application of artificial neural networks to rainfall forecasting in Queensland, Australia," *Adv. Atmosp. Sci.*, vol. 29, no. 4, pp. 717–730, Jan. 2012.

[18] A. Haidar and B. Verma, "Monthly rainfall forecasting using one-dimensional deep convolutional neural network," *IEEE Access*, vol. 6, pp. 69053–69063, 2018.

[19] B. Athiwaratkun and J. W. Stokes, "Malware classification with LSTM and GRU language models and a character-level CNN," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, New Orleans, LA, USA, 2017, pp. 2482–2486.

[20] G. F. Lin and M. C. Wu, "A hybrid neural network model for typhoon-rainfall forecasting," *J. Hydrol.*, vol. 375, nos. 3–4, pp. 450–458, Sep. 2009.

[21] P. Zhang, Y. Jia, L. Zhang, J. Gao, and H. Leung, "A deep belief network based precipitation forecast approach using multiple environmental factors," *Intell. Data Anal.*, vol. 22, no. 4, pp. 843–866, Jun. 2018.

[22] F. Mekanik, M. Imteaz, S. Gato-Trinidad, and A. Elmahdi, "Multiple regression and artificial neural network for long-term rainfall forecasting using large scale climate modes," *J. Hydrol.*, vol. 503, pp. 11–21, Oct. 2013.

[23] X. Shi *et al.*, "Deep learning for precipitation nowcasting: A benchmark and a new model," in *Advances in Neural Information Processing Systems*. Red Hook, NY, USA: Curran Assoc., Inc., Dec. 2017, pp. 5622–5632.

[24] *Meteorological Verification Data Technical*, Bureau Meteorol., Melbourne, VIC, Australia, 2015. Accessed: Feb. 28, 2020. [Online]. Available: https://data.gov.au/search?q=Rainfall%20and%20temperature%20forecast%20and%20observations
[25] J. Yang, K. Shen, C. Ong, and X. Li, "Feature selection for MLP neural network: The use of random permutation of probabilistic outputs," *IEEE Trans. Neural Netw.*, vol. 20, no. 12, pp. 1911–1922, Dec. 2009.
[26] J. Yang and C. Ong, "Feature selection using probabilistic prediction of support vector regression," *IEEE Trans. Neural Netw.*, vol. 22, no. 6, pp. 954–962, Jun. 2011.
[27] A. Graves, N. Jaitly, and A. Mohamed, "Hybrid speech recognition with deep bidirectional LSTM," in *Proc. IEEE Workshop Autom. Speech Recognit. Understand.*, Olomouc, Czech Republic, Dec. 2013, pp. 273–278.
[28] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, no. 1, pp. 281–305, Feb. 2012.
[29] Y. Wang, Y. Shen, S. Mao, X. Chen, and H. Zou, "LASSO and LSTM integrated temporal model for short-term solar intensity forecasting," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2933–2944, Apr. 2019.

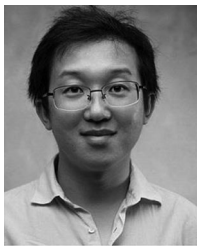
Neethu Madhukumar (Student Member, IEEE) received the B.Tech. and M.Tech. degrees in electronics and communication engineering from the University of Kerala, Thiruvananthapuram, India, in 2012 and 2015, respectively. She is currently pursuing the Ph.D. degree with the College of Science and Engineering, James Cook University, Cairns, QLD, Australia.

She was involved in teaching the subject random processes and applications for the undergraduate students with the College of Engineering Trivandrum as part of the Govt. of Kerala project "Step4U" for a year. As part of her master's research, she has worked with the National MEMS Design Lab, College of Engineering, Trivandrum, and built a structural health monitoring system using MEMS-based wireless sensors. Her research interests include time-series modeling, wireless sensor networks, deep learning, and the Internet of Things.

Eric (Gengkun) Wang (Member, IEEE) received the B.Eng. degree in mechanical engineering and the M.Eng. degree in mechatronic engineering from the University of Science and Technology, Beijing, China, in 2006 and 2009, respectively, and the Ph.D. degree in telecommunications engineering from the University of Southern Queensland (USQ), Toowoomba, QLD, Australia, in 2013.

He has been working as a Postdoctoral Research Fellow with the School of Electrical and Mechanical Engineering, USQ, for two and a half years. He has authored more than 20 high-quality papers and participated in many national and international research projects.

Dr. Wang received the Best Paper Award from the IEEE Wireless Communications and Networking Conference, Cancun, Mexico, in 2011.



Yi-Fan Zhang (Member, IEEE) received the B.Eng. and M.Eng. degrees in reliability and systems engineering from Beijing University of Aeronautics and Astronautics, Beijing, China, in 2008 and 2011, respectively, and the Ph.D. degree in data science from the Queensland University of Technology, Brisbane, QLD, Australia, in 2016.

From 2017 to 2020, he had worked as a Postdoctoral Fellow with CSIRO, Agriculture and Food, Canberra, ACT, Australia. Since 2020, he has been a Postdoctoral Fellow with the Queensland Alliance for Agriculture and Food Innovation, University of Queensland, Brisbane. His research interests include artificial intelligence, time-series modeling, image processing, and the Internet of Things.

Wei Xiang (Senior Member, IEEE) received the B.Eng. and M.Eng. degrees in electronic engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 1997 and 2000, respectively, and the Ph.D. degree in telecommunications engineering from the University of South Australia, Adelaide, SA, Australia, in 2004.

He is the Cisco Chair of AI and Internet of Things with La Trobe University, Melbourne, VIC, Australia. He was the Foundation Chair and the Head of the Discipline of Internet of Things Engineering, James Cook University, Cairns, QLD, Australia. Due to his instrumental leadership in establishing Australia's first accredited Internet of Things Engineering degree program, he was inducted into Pearcey Foundation's Hall of Fame in October 2018. He has published over 250 peer-reviewed papers, including three academic books and 180 journal articles. His research interests include Internet of Things, wireless communications, machine learning for IoT data analytics, and computer vision.

Dr. Xiang received the TNQ Innovation Award in 2016, the Pearcey Entrepreneurship Award in 2017, and the Engineers Australia Cairns Engineer of the Year in 2017. He was a co-recipient of four Best Paper Awards at WiSATS'2019, WCSP'2015, IEEE WCNC'2011, and ICWMC'2009. He has been awarded several prestigious fellowship titles. He was named a Queensland International Fellow from 2010 to 2011 by the Queensland Government of Australia, an Endeavour Research Fellow from 2012 to 2013 by the Commonwealth Government of Australia, a Smart Futures Fellow from 2012 to 2015 by the Queensland Government of Australia, and a JSPS Invitational Fellow jointly by the Australian Academy of Science and Japanese Society for Promotion of Science from 2014 to 2015. He is the Vice Chair of the IEEE Northern Australia Section. He was an Editor for IEEE COMMUNICATIONS LETTERS from 2015 to 2017, and is an Associate Editor for IEEE ACCESS and *Telecommunications Systems* (Springer). He has severed in a large number of international conferences in the capacity of the general co-chair, the TPC co-chair, and the symposium chair. He is an elected Fellow of IET in U.K. and Engineers Australia.