Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 135 (2018) 89–98

Procedia Computer Science

www.elsevier.com/locate/procedia

3rd International Conference on Computer Science and Computational Intelligence 2018

Single Layer & Multi-layer Long Short-Term Memory (LSTM) Model with Intermediate Variables for Weather Forecasting

Afan Galih Salman[a*], Yaya Heryadi[b], Edi Abdurahman[b], Wayan Suparta[c]

[a]*Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480.*
[a,b]*Computer Science Department, BINUS Graduate Program-Doctor of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480,*
[c]*Department of Electrical Engineering, Faculty of Science and Technology, Sanata Dharma University, Yogyakarta 55282,*
* *asalman@binus.edu*

Abstract

Weather forecasting has gained attention many researchers from various research communities due to its effect to the global human life. The emerging deep learning techniques in the last decade coupled and the wide availability of massive weather observation data have motivated many researches to explore hidden hierarchical pattern in the large volume of weather dataset for weather forecasting.
The purposes of this research are to build a robust and adaptive statistical model for forecasting univariate weather variable in Indonesian airport area and to explore the effect of intermediate weather variable related to accuracy prediction using single layer Long Short Memory Model (LSTM) model and multi layers LSTM model. The proposed forecasting model is an extension of LSTM model by adding intermediate variable signal into LSTM memory block. The premise is that two highly related patterns in input dataset will rectify the input patterns so make it easier for the model to learn and recognize the pattern from the training dataset. In an effort to achieve a robust model for learning and recognizing weather pattern, this research will also explore various architectures such as single layer LSTM and Multiple Layer LSTM (4 layers LSTM). The dataset is weather variable data collected by Weather Underground at Hang Nadim Indonesia airport. This research used visibility as predicted data and temperature, pressure, humidity, dew point as intermediates data. The best model of LSTM in this experiment is multiple layers LSTM and the best intermediate data is pressure variable. Using the pressure variable this model has gained the validation accuracy 0.8060 and RMSE 0.0775.

Keywords: *weather forecasting, LSTM, intermediate, validation accuracy, RMSE*

1. Introduction

1877-0509 © 2018 The Authors. Published by Elsevier Ltd.

10.1016/j.procs.2018.08.153

Weather forecasting is an interesting research problem in flight navigation area. One of the important weather data in aviation is visibility. Visibility is an important factor in all phases of flight, but especially when the aircraft take-off and initial climb, approach and landing, and taxi-in.

Koetse show the impact of climate change and weather on transport in San Francisco International Airport[1]. A study by Eads shows that poor visibility in the summer months and rain storms in the winter months lead to substantial delays and numerous cancellations[2].

In the last decade, there are vsome research on weather forecasting using Neural Network (NN) which can be found in literature. NN model for predicting rainfall[3], Recurrent Neural Network (RNN) model for predicting regional annual runoff[4], Fuzzy time series model for temperature prediction[5] , Ensemble of NNs model for predicting temperature, wind speed and humidity[6] , Chaotic Oscillatory-based NN for short term wind forecasting using LIDAR Data[7], NN fuzzy wavelet model for long term rainfall forecasting[8].

This study propose a model to analize the relationship between two influencing weather variables in weather prediction, for example the visibility is influenced by dew point or humidity as intermediate variables. This propose model is using single and multi layer Long Short Memory Model (LSTM).

LSTM is a recurrence Neural Networks proposed by Hochreiter and Schmidhuber. LSTM is a specific recurrent neural network (RNN) architecture that designed a model temporal sequences with their long-range dependencies[9] .

The purposes of this research are to build a robust and adaptive statistical model for forecasting univariate weather variable in Indonesian airport area and to explore the effect of intermediate weather variable related to accuracy prediction of single layer LSTM model and multi layers LSTM model.

2. Literature Review

In the last decade, many significant efforts to solve weather forecasting problem using statistical modeling including machine learning techniques have been reported. Xingxian proposed LSTM with the Trajectory GRU (TrajGRU) model to predict the future rainfall intensity in a local region over a relatively short period of time[10]. Seongchan Kim proposed model to predicts the amount of rainfall from weather radar data using convolutional LSTM (ConvLSTM). ConvLSTM is a variant of LSTM (Long Short-Term Memory) containing a convolution operation inside the LSTM cell[11].

Isabelle Roesch developed a recurrent convolutional neural network to forecast meteorological attributes, such as temperature, pressure and wind speed. Isabelle present a visualization system to helped user quickly assess, adjust and improve the network design[12]. Aditya Grover developed a hybrid approach model that combines the distinctive trained predictive models and a deep neural network. This models make the joint statistics that consist a set of weather-related variables[13].

Bedaiko developed LSTM using various complex networks metrics to forecast ENSO phenomenon[14]. The preliminary experiments show that with more data sample and a quite complex LSTM neural network model can make a great potential for forecasting ENSO phenomenon.

Recurrent Neural Networks (RNN)

Recurrent neural network (RNN) is an artificial Neural Network (NN) use for time series prediction. Elman networks is a class of RNN consists of one or more hidden layer. The first layer has the weight that is obtained from the input layer. Every layer will receive weight from the previous layer. This network usually uses the activation function of sigmoid bipolar for the hidden layer and linear function (purelin) for the output layer. This Elman network has activation function both continue and discontinue. Delay that is happened in this connection between the input layer and the first hidden layer in the previous time (t-1) that can be used in the current time (t). The unique of the recurrent neural network is the feedback connection which conveys interference information (noise) at the previous input that will be accommodated to the next input (see figure 1).

Let $x(t)$ and $y(t)$ be input and output time series respectively; the three connection weight matrices are W_{IH}, W_{HH}, and W_{HO}; hidden and output unit activation functions are f_H and f_O; the behavior of the recurrent network can be described by the pair of non-linear matrix equations:

$$h(t+1) = f_H\big(W_{IH}x(t) + W_{HH}h(t)\big) \tag{1}$$

$$y(t+1) = f_O\big(W_{HO}h(t+1)\big). \tag{2}$$

Where: $h(t)$ represent a state of a dynamical system. $h(t)$is a set of values that summarizes all the information about the past behaviour of the system that is necessary to provide a description of its future behavior.[15]

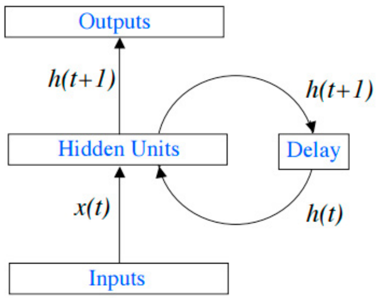


Fig.1. Recurrent Neural Network Model

Long Short-Term Memory (LSTM)

LSTM is a specific recurrent neural network (RNN) architecture that was designed to model temporal sequences. LSTM has a long-range dependencies that make LSTM more accurately than conventional RNNs. Backpropagation algorithm in RNN architecture causes error backflow problem[9].

Unlike RNN, LSTM contains special units called memory blocks in the recurrent hidden layer. The memory blocks contain memory cells with self-connections storing the temporal state of the network in addition to special multiplicative units called gates to control the flow of information. Each memory block in the original architecture contained three gate types which are namely:

- Input gate: the input gate controls the flow of input activations into the memory cell.
- Output gate: output gate controls the output flow of cell activations into the rest of the network.
- Forget gate: scales the internal state of the cell before adding it as input to the cell through the self-recurrent connection of the cell, therefore adaptively forgetting or resetting the cell's memory.

In addition, the modern LSTM architecture contains peephole connections from its internal cells to the gates in the same cell to learn precise timing of the output.

In order to make analysis easier, LSTM architecture is often unfolded over t (time)-dimension which can be represented by the following diagram (see figure 2).

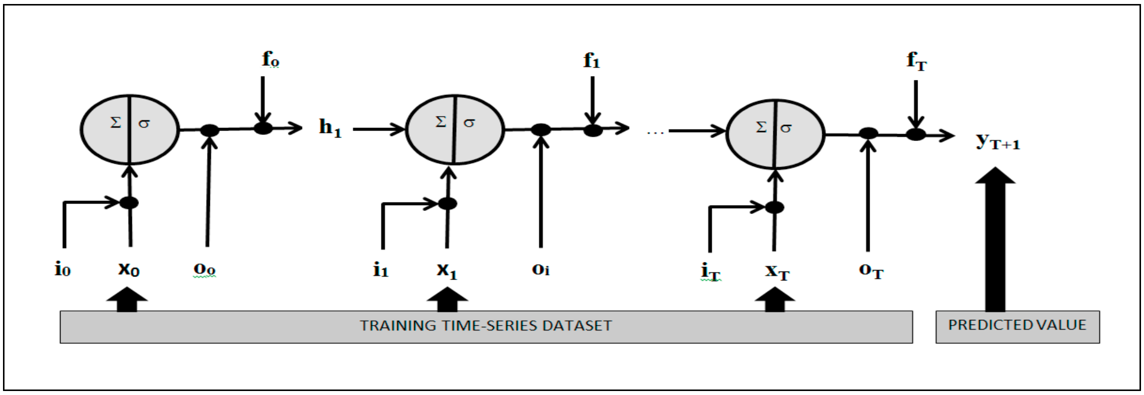


Fig.2. Architecture of Unfolded Long Short-Term Memory Model.

From the diagram in figure 3, it can be seen that each LSTM block receives the following signals: input signal (x), input gate signal (i), recurrent signal (h), and forget gate signal (f); and produces output gate signal (o).

The flow of process in each LSTM memory block can be represented by the following diagram.

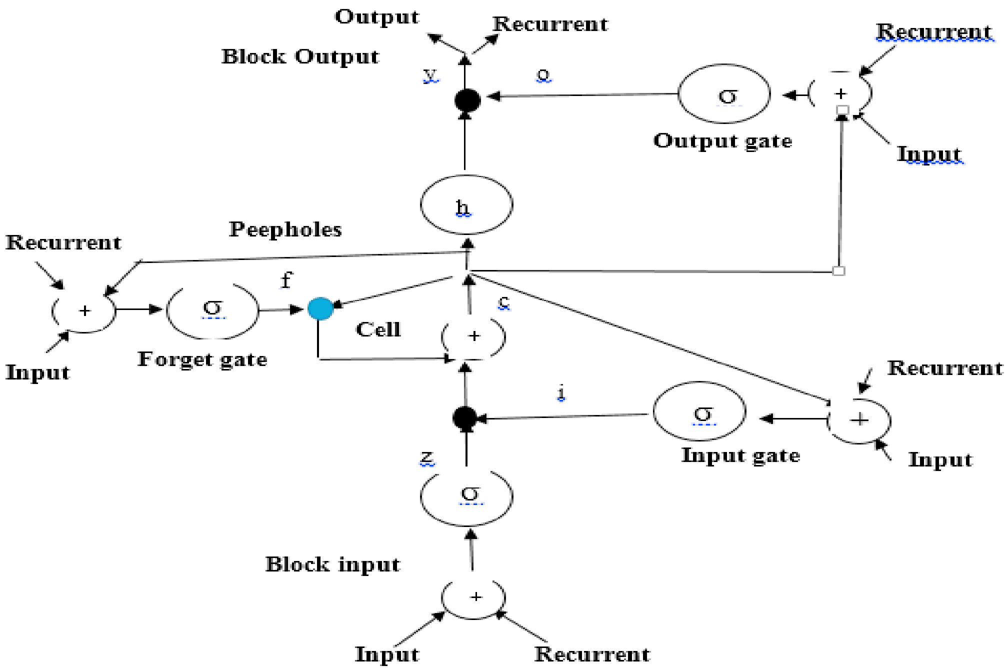


Fig.3. Computation Flow Diagram of Long Short-Term Memory Model

An LSTM network computes a mapping from an input sequence $x = (x_1, \ldots, x_T)$ to an output sequence $y = (y_1, \ldots, y_T)$ by calculating the network unit activations using the following equations iteratively from $t = 1$ to T as follows.

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{3}$$

$$z_t = tanh(W_z x_t + U_z h_{t-1} + b_z) \tag{4}$$

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \tag{5}$$

$$C_t = i_t * z_t + f_t * C_{t-1} \tag{6}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + V_o C_t + b_o) \quad (7)$$

$$h_t = o_t * tanh(C_t) \quad (8)$$

Where: $W_i, W_z, W_f, W_o, U_i, U_z, U_f, U_o$ are model parameters to be estimated during model training; σ (sigmoid) and $tanh$ are activation functions and$b$'s are biases.

3. Dataset and Data Preprocessing

Dataset for this research was obtained from Weather Underground (https://www.wunderground.com/) which collects weather data including temperature, dew point, humidity and visibility from many weather stations all over the world. The range of data for this study was from year 2012 to year 2016 comprise of 40,025 time series data at Hang Nadim Airport Indonesia.

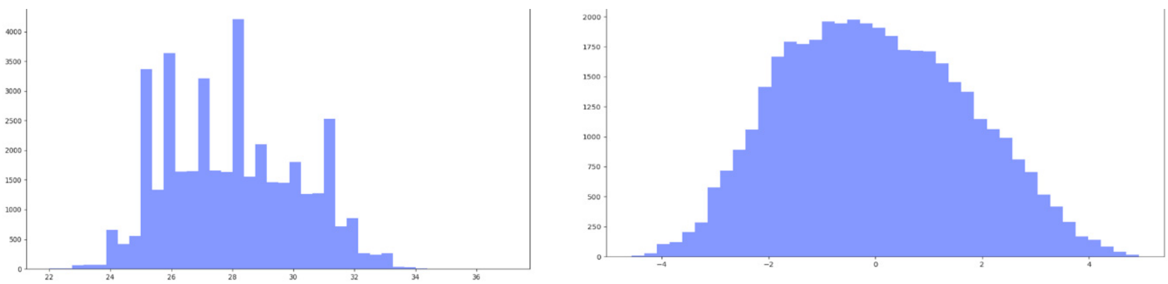
The main data preprocessings applied to raw visibility timeseries data are: normalization (9), rescaling into range [0,1] (10) and smoothing using moving average (MA) with lag = 9 (11). Consider weather time series data in T time interval: $X = [x_1, x_2, \ldots, x_T]$

$$x_t = \frac{x_t - \bar{\bar{x}}}{s_x} \quad (9)$$

$$x_t' = \frac{x_t - xmin}{xmax - xmin} \quad (10)$$

$$x_t'' = \frac{1}{9}(x_t' + x_{t-1}' + \cdots + x_{t-8}') \quad (11)$$

Where: x_t is observation at t, x_t' is normalized data at t, and x_t'' is the result of data smoothing using moving average at t. A sample of histograms of raw data and smoothed data using moving average are shown in Figure 4. From Figure.4 it appears that the raw data temperature distribution is rather skewed (a) and after being pre-processed, the data temperature distribution looked a bit smoother (b).



(a) (b)

Fig 4. Temperature Data Distribution: (a) Raw Data, and (b) After Pre-processed

4. Method

Research Frameworks for this study can be describe using the following diagram in figure 5. As can be seen from Figure 5, the proposed model is a stacked LSTM with subsequent layers having 200, 100, 90, and 50 nodes of hidden layers. The last part of the model is a fully connected neural network with 1 output nodes. The function activation is sigmoid function and the model was set for 500 Epochs. Validation splits are 70% data for training and 30% data for testing. This model is using early stopping algorithm to gain the best value of validation accuracy and RMSE.

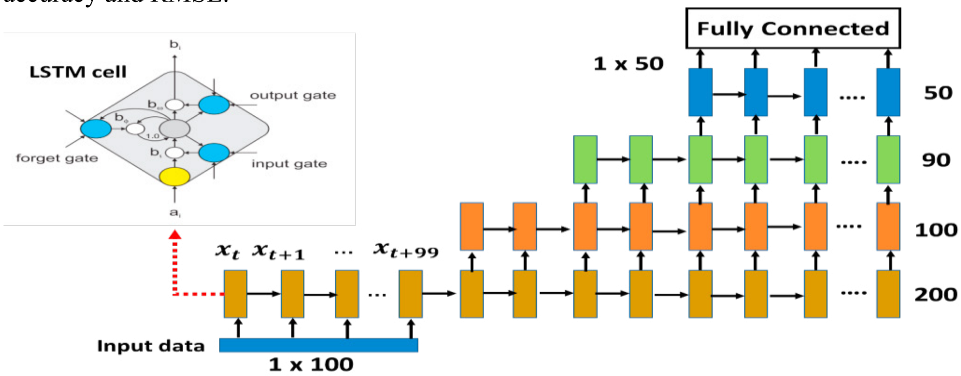


Fig.5. Research Framework

LSTM's memory cell is a basic unit of LSTM model whose structure can be illustrated using Figure 6. As described by each memory cell contains input gate that learns to protect the constant error flow within the memory cell from irrelevant inputs. Output gate unit learns to protect other units from irrelevant memory contents stored in the memory cell. Forget gate unit learns to control the extent to which a value remains in the memory cell.

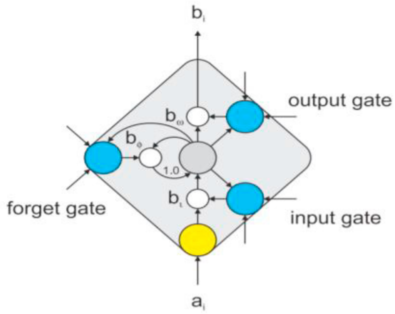


Fig.6. Structure of LSTM cell

For the example, input data consist predict variable weather data such as visibility and intermediate data variable such as humidity and dew point. Output is predicted variable data visibility.

In this experiment used software such as Python program language, Tensor Flow, KERAS, Ubuntu and hardware with specifications such as Intel Core i5-4200 U CPU@1,60 GHz, 64-bit Operating System and RAM 4GB.

5. Evaluation

After some pre experiment, this study has obtained selected variation of lookback and numbers of prediction. They are 100 numbers of lookback and 2 numbers of prediction nodes.

The next experiment is implementation of single layer LSTM model & multi layers LSTM model in order to predict visibility with varian intermediate variables such as humidity, dew point, temperature and pressure.

The implementation of single layer LSTM and multilayer LSTM resulted in the value of validation accuracy.

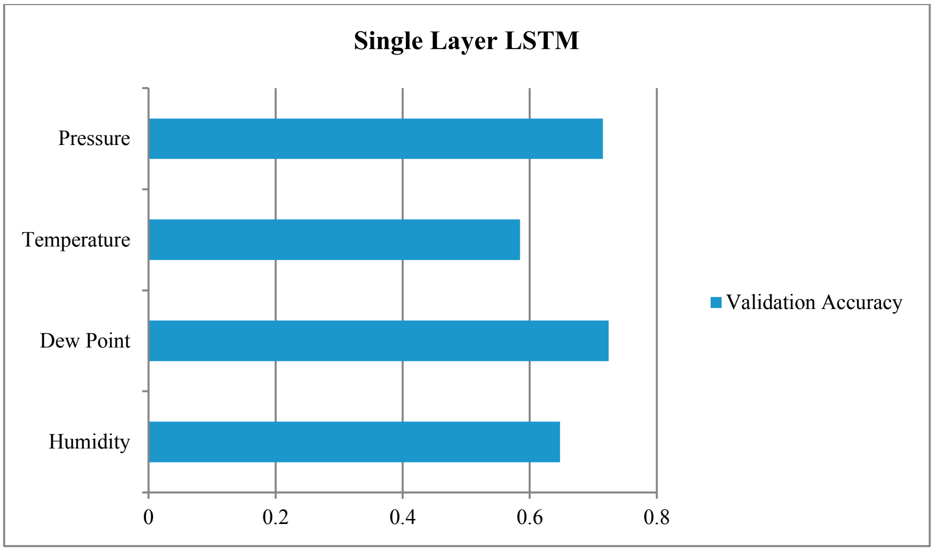


Fig.7. Validation accuracy of single layer LSTM

Figure 7 describes validation accuracy for every intermediate variable using single layer LSTM. The figure 7 show that intermediate variable pressure and dew point has big impact on accuracy of visibility prediction which are more than 70 percent. The best validation accuracy of Intermediate variable is 0.7243 achieved by dew point.

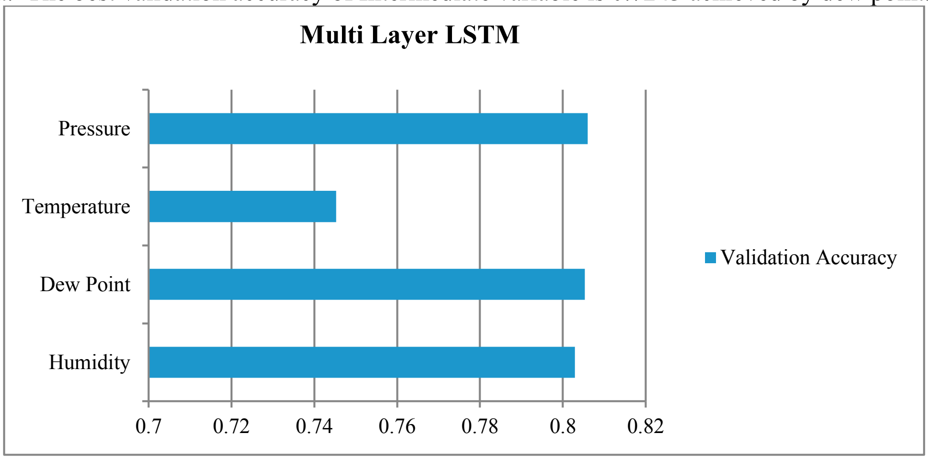


Fig.8. Validation accuracy of multi layer LSTM

Figure 8 describes validation accuracy for every intermediate variable using multi layer LSTM. Three intermediate variable that are pressure, dew point, and humidity achieve accuracy of visibility prediction more than 80 percent. The best validation accuracy is 0.8060 achieved by pressure variable.

In fact, the multi layer LSTM better than single layer LSTM as desribed by the table 1 below that show the all best values of validation accuracy achieved by intermediate variable on multi layer LSTM.

Table.1. The highest validation accuracy

Intermediate Variable	Highest Validation Accuracy	RMSE	LSTM Model
Humidity	0.8029	0.0771	Multi-layer
Dew point	0.8053	0.0778	Multi-layer
Temperature	0.7453	0.0822	Multi-layer
Pressure	0.8060	0.0775	Multi-layer

As described in table 1, the best intermediate data in Hang Nadim Airport is pressure variable that resulted in validation accuracy 0.8060 with RMSE 0.0775. The second best intermediate data is dew point variable reaching validation accuracy 0.8053 and RMSE value 0.0778. The third best intermediate data is humidity variable achieving validation accuracy 0.8029 and RMSE value 0.0771.Tabel 1 show that multi layer LSTM resulted in better accuracy prediction of visibility than single layer LSTM.

The correlation between RMSE and numbers of epochs described in figure 9.

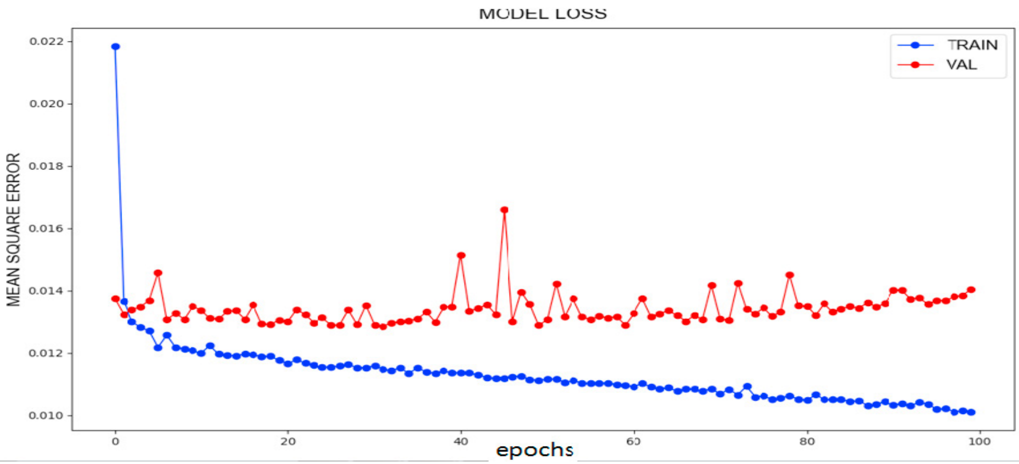


Fig.9. The chart of model loss for intermediate variable dew point

Figure 9 show the graph of RMSE for intermediate variable pressure show that train accuracy decrease along with the addition number of epoch and RMSE value of validation accuracy increase along the addition number of epochs. The best value RMSE of validation accuracy is 0.0775 in range100 epochs. The pressure variable resulted in relatively low RMSE compare with three other variables.

Correlation between validation and training for data intermediate pressure is described in figure 10.

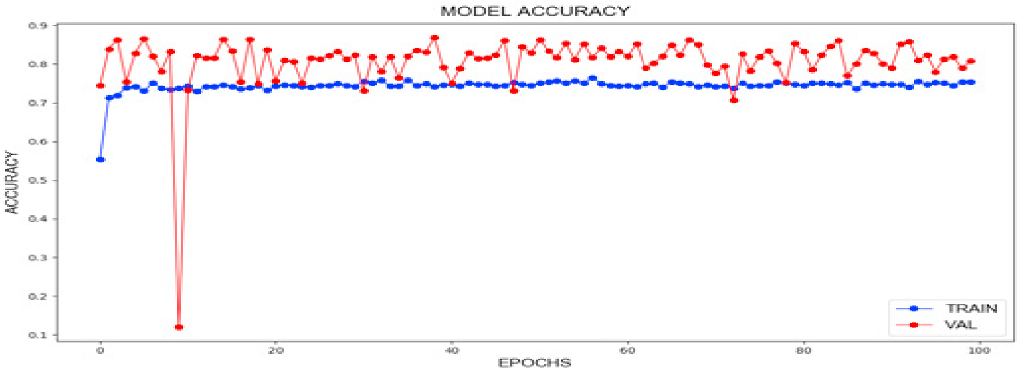


Fig.10. The chart of model accuracy for intermediate data pressure

Figure 10 show that accuracy of the train and validation intermediate data pressure is consistent after 20 epoch. The best validation accuracy is 0.8060 in range100 epochs. Compare with three other intermediate variables, the pressure variable resulted in the best validation accuracy.

This experiment show that LSTM model is able to explain or formulate relationship among the predicted and intermediate variable. The addition of intermediate variables able to increase accuration of weather prediction. In this experiment to predict visibility with addition intermediate variables pressure produce the best accuration and the lowest RMSE using multi layer LSTM.

These research try to modify of input weather data that has influence each other to find the combination weather data input that can optimize forecasting accuracy in time series data model. The combination of input weather data model which can be used for weather forecasting in Airport area.

6. Conclusion

Weather forecasting task has gained wide attention from many research communities due to its significant effect to global human life. Many efforts to build weather forecasting models have been proposed resulted in a vast number of publications available in literature. However, the nature of weather is so complex that impossible to be formulated in a single mathematical model.

Despite many models have been proposed for weather prediction, most of these models used the same input and output variables. The result of this study, which exploited LSTM model variant, showed that intermediate variables can improve prediction capability of the model

The LSTM model is feasible and suggested to be implemented in predicting weather with the addition of intermediate data in order to improve the accuration. The best model of LSTM model in this experiment is multiple layers LSTM and the best intermediate data is pressure variable. Using the pressure variable this model has gained the validation accuracy 0.8060 and RMSE 0,0775.

The most important findings of these research are :

- The combination of predicted variable and intermediate variables can optimize forecasting accuracy in time series data model.
- The research artifacts (scripts and dataset) will be available for other researchers in the same domain.

The suggestion for further research is trying other algorithms for smoothing data input and implementing the combination of LSTM and Convolutional Neural Network (CNN) models.

References

[1] Koetse. The impact of climate change and weather on transport: An overview of empirical findings. Transportation Research Part D 14 Elsevier; 2009, p 205–221

[2] Eads, G.C., Kiefer, M., Mendiratta, S., McKnight, P., Laing, E., Kemp, M.A.. Reducing weather-related delays and cancellations at San Francisco International Airport. CRA Report No. D01868-00. Prepared for San Francisco International Airport, Charles River Associates, Boston;2000.

[3] Normakristagaluh P. Artificial Neural Network for Rainfall forecasting in statistical downscaling. Institut Pertanian Bogor, Thesis; 2004.
[4] Afan Galih Salman, Bayu Kanigoro, Yaya Heryadi, Weather forecasting using deep learning techiniques, ICACSIS UI Proceeding IEEE; 2015.
[5] Chen, S.-M., and J.-R. Hwang. Temperature prediction using fuzzy time series. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 30.2; 2000, p. 263-275.
[6] Maqsood, I., M. R. Khan, and A. Abraham. An ensemble of neural networks for weather forecasting. Neural Computing & Applications 13.2.; 2004, p 112-122.
[7] Kwong, K. M., Liu, J. N. K., Chan, P. W., and Lee, R. Using LIDAR Doppler velocity data and chaotic oscillatory-based neural network for the forecast of meso-scale wind field, In Evolutionary Computation, CEC 2008;2008. p. 2012-2019
[8] Afshin. Long term rainfall forecasting by integrated artificial ,neural network-fuzzy logic-wavelet model in Karoon Basin, Scientific Research and Essay vol 6(6); 2011, pp.1200-1208
[9] Hochreiter, S. and J. Schmidhuber, Long short term memory, Neural Computation MIT; 1997.
[10]Xingjian Shi Zhourong Chen Hao Wang Dit-Yan Yeung, Convolutional LSTM network: a machine learning approach for precipitation nowcasting, Part of: Advances in Neural Information Processing Systems 28 (NIPS); 2015.
[11]Seongchan Kim, Seungkyun Hong, Minsu Joh, Sa-kwang Song, Deep Rain: ConvLSTM network for precipitation prediction using multichannel radar data. 7th International Workshop On Climate Informatics, Sept 20-22; 2017.
[12]Isabelle Roesch and Tobias Günther, Visualization of neural network predictions for weather forecasting. Eurographics Proceedings, The Eurographics Association; 2017.
[13]Aditya Grover,, Ashish Kapoor,, Eric Horvitz, A deep hybrid model for weather forecasting, KDD '15 Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '15 Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2015, p 379-386.
[14]Clifford Broni-Bedaiko, Ferdinand Apietu Katsriku, Tatsuo Unemi, Norihiko Shinomiya, Jamal-Deen Abdulai and Masayasu Atsumi. El niño-southern oscillation forecasting using complex networks analysis of LSTM networks. The Twenty-Third International Symposium on Artificial Life and Robotic; 2018.
[15] Rumelhart, D., G. Hinton, and R. Williams. Learning internal representations by error propagation. In Parallel Dist. Proc. MIT Press; 1986, pp. 318-362.