ICIC Express Letters ICIC International ©2020 ISSN 1881-803X
Volume **14**, Number **10**, October **2020** pp. **943–949**

ENHANCED LSTM MULTIVARIATE TIME SERIES FORECASTING FOR CROP PEST ATTACK PREDICTION

Teguh Wahyono[1,2], Yaya Heryadi[1], Haryono Soeparno[3] and Bahtiar Saleh Abbas[4]

[1]Doctor of Computer Science, Binus Graduate Program
[3]Bioinformatics and Data Science Research Center
[4]Department of Industrial Engineering
Bina Nusantara University
Jl. K. H. Syahdan No. 9, Kemanggisan, Palmerah, Jakarta 11480, Indonesia
{ teguh.wahyono; yayaheryadi; haryono; bahtiar }@binus.edu

[2]Faculty of Information Technology
Satya Wacana Christian University
Jl. Diponegoro 52-60, Salatiga 50711, Indonesia
teguh.wahyono@uksw.edu

Received December 2019; accepted March 2020

Abstract. *This paper proposes the implementation of Bidirectional Long Short-Term Memory (LSTM) model for forecasting crop pest attack based on multivariate inputs of weather data. To enhance the model performance, the efforts implemented include optimizing the training process by applying sliding-windows methods and setting appropriate parameters. After obtaining the best settings, the proposed model was then compared with two other variants of the LSTM model, namely Vanilla LSTM and Stack LSTM. The model performance evaluation was done by finding Mean Square Error (MSE) and root mean square error values from training and testing process. The results show that there were increased performances from using sliding-windows method to optimize the training process and the appropriate parameter setting. Bidirectional LSTM model, along with the treatments implemented, gave the best performance compared to the two other models.*
Keywords: LSTM, Multivariate, Time series forecasting, Pest attack

1. **Introduction.** Pest attack is one of major factors that cause crop failures in Indonesia, besides drought and natural disasters [1]. During the harvesting period of 2016/2017 for instance, approximately 63,000 ha of rice fields or almost 0.5% of the total farm lands in Indonesia were attacked by pests and experienced drought [2]. From 63,075 ha of the fields attacked by pests, 20,152 ha experienced crop failures. The attack is fluctuating from year to year and causes the decrease of farming productivity in Indonesia. One of the efforts done to anticipate the pest attack is developing a system to predict the attack, so that the loss caused by the attack can be minimized.

Many studies on crop pest attack prediction have been done by researchers. Most of the studies were multivariate time-series forecasting cases as they used many variables as input data. Various methods from linear, clustering to neural networks had also been implemented to get the optimal results [1,3]. A classic linear method was used by Wu et al. [4] to predict brown plant hopper pest by using the climate data as the input. On the other hand, Hooker et al. [5] used multiple linear regression method to predict the pest of wheat plants by using rainfall and temperature data. Space Time Autoregressive Integrated Moving Average (STARIMA) and Generalized Space Time Autoregressive Integrated Moving Average (GSTARIMA) methods, which were the variants of Vector Autoregressive Integrated Moving Average (VARIMA), were used to forecast scirpophaga innotata

DOI: 10.24507/icicel.14.10.943

pest attack [6]. Meanwhile, Gaussian Naïve Bayes method was used to predict pest attack through data mining technique [7] and hemileia vastatrix disease in coffee plants [8].

The classic time-series methods employed in the above examples had several weaknesses, including relatively strict requirements, linear dependency from time to time, inflexibility in scalability and high dimensional data, as well as incompatibility to long-term dependencies model. A more flexible approach to modeling multivariable relationships was to use the neural network model. For example, studies done by Klem et al. [9] predicted deoxynivalenol disease in wheat plants. Also, Wibowo and Sutikno [10] used neural network to predict pest attack in paddy plants. Neural network model is considered fairly flexible as it does not require data assumptions that are often difficult to fulfill [11].

One of neural network architectures considered to be successful in modeling non-linear temporary data is Long Short-Term Memory (LSTM) neural network. Kim and Chung [12] used LSTM to predict pests in cotton, while Chen et al. [13] applied the method for early forecasting in rice blast disease. Those studies show that LSTM had good performance in multivariate time-series forecasting. The model can also handle long-term dependencies that often become an issue in classic time-series methods [11]. Other studies done by Maholtra et al. [14] and AlDosari [15] also indicate that LSTM has the ability to work effectively in many studies on forecasting. Nevertheless, some of these studies [14,15] recommend further research to improve its prediction results, especially in terms of optimizing the training process as will be discussed in this paper.

This research proposes the implementation of LSTM multivariate forecasting time-series with bidirectional model in a case of paddy or rice crop pests attack by using climate data as the input variable. Boyolali Regency of Central Java Province was chosen as the object of the research since the region was one of rice production centers in Indonesia that had many pest attack cases [1]. Several experiments were done to enhance the performance of the model in predicting. Some efforts implemented include optimizing the training process by applying sliding-windows methods and setting appropriate parameter so that the best performance could be achieved. Sliding windows is a way to transform data so that the input data can be used appropriately in LSTM model [16,17]. An appropriate setting of the window can minimize the loss function of the model [17]. Finally, the proposed model was compared with two other variants of the LSTM model namely Vanilla LSTM and Stack LSTM.

The rest of the paper is organized as follows. Part two discusses about the research methodology and the algorithm used in this research. Part three describes the experiment result of the research and the last part consists of conclusions and future work.

2. Methodology.

2.1. Proposed method. This paper proposes the implementation of Bidirectional Long Short-Term Memory (LSTM) model to forecast crop pest attack. The proposed method is shown in the diagram in Figure 1. The input data used were climate data (including temperature, rainfall, and humidity) and area coverage of pest attack arranged in time-series.

Before they were used for training models, a pre-processing phase was done by normalizing and transforming the data. The normalization of the data was carried out by using min-max scaler, whereas the data transformation was done by using sliding-windows method. Sliding windows is a way to transform data based on specific time frame so that the input data can be used appropriately in LSTM model [16,17]. For example, if the current time is t and the predicted value is at $t+1$, then the input variables are t, $t-1$, $t-2$ to $t-n$. The width of the data, n, is called window [16].

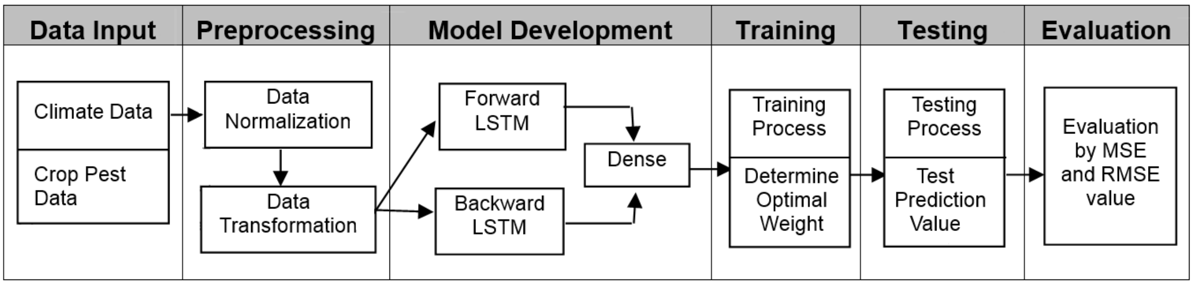

Figure 1. Proposed method design

The proposed model was Bidirectional LSTM which combined Forward LSTM and Backward LSTM densely connected as an additional layer. The next step was to process training and testing by using the developed model. To optimize the training process, an appropriate parameter adjustment was performed to obtain optimal results. These parameters include epoch and batch size to get an optimal learning rate. Experimental approach is used to find the best parameter values for the given data set.

The model performance evaluation was done by finding Mean Square Error (MSE) and root mean square error values from training and testing process. MSE was used to measure the consistency of the model with square differences between actual data and predicted data, as shown in the equation below.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(X_{obs,i} - X_{pred,i})^2 \tag{1}$$

where n is number of data, X_{obs} is observed values and X_{pred} is the predicted values.

Furthermore, RMSE was used to measure the accuracy level of the model's prediction results. The value of the errors generated by the prediction model could be seen from the RMSE value. The RMSE calculation can be seen in Equation (2).

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(X_{obs,i} - X_{pred,i})^2}{n}} \tag{2}$$

where n is number of data, X_{obs} is observed values and X_{pred} is the predicted values. The lower RMSE means the performance of the model is better. It means that the value of the prediction was close to the value in the reality. Finally, the results of those three models, Vanilla LSTM, Stack LSTM and Bidirectional LSTM, were compared to find the optimal model.

2.2. **Dataset.** This study used a dataset consisting of climate data and area coverage data of scirpophaga innotata pest attack in Boyolali Regency of Central Java Province, Indonesia. The dataset was arranged in multivariate time-series. The climate data used in this study were average daily temperature, humidity and rainfall during 15 years period from 2001 to 2016. The daily rainfall data and the pest attack data were gathered from daily observations done by The Center of Food Crops and Horticulture Protection of Central Java Province. Meanwhile, the data of humidity and the length of sunray exposure were obtained from Weather Observation Station of Indonesia Air Force in Adisumarmo Airport, Solo.

2.3. **LSTM model.** LSTM is one of deep learning methods, which is the development of Recurrent Neural Network (RNN) known to have a good ability in many studies on time-series forecasting and to solve problems in long-term dependencies model [14,15].

As Figure 2 indicates, an LSTM network has three main gateways, which are input gate, forget gate, and output gate [18]. In this model, the input units are denoted by x,

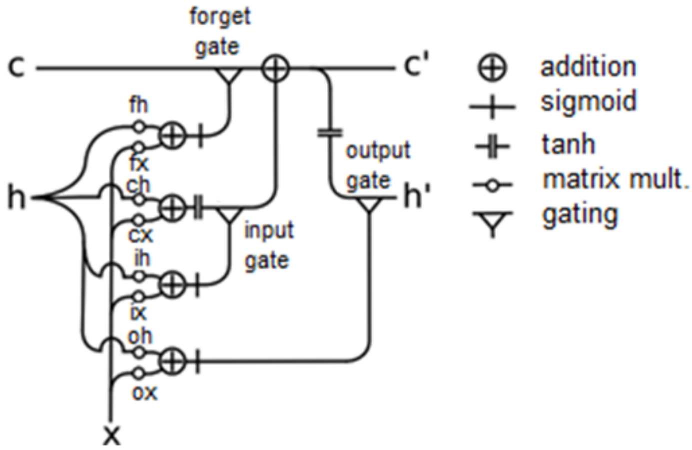

FIGURE 2. LSTM architecture

the output units by y and the hidden units by h. Input gate (i) will determine the new information stored in the cell state with the formula seen in Equation (3).

$$\vec{i}_t = \sigma\left(\vec{x}_t W_{xi} + \vec{h}_{t-1} W_{hi} + \vec{c}_{t-1} W_{ci} + \vec{b}_i\right) \tag{3}$$

where t is discrete time ($t = 1, 2, \ldots, n$), σ is the logistic sigmoid activation function, W is the weights, c is memory cell units and b is the bias. Forget gate (f) will determine the information that should be deleted from a cell state to diminish dependency toward long-term memory.

$$\vec{f}_t = \sigma\left(\vec{x}_t W_{xf} + \vec{h}_{t-1} W_{hf} + \vec{c}_{t-1} W_{cf} + \vec{b}_f\right) \tag{4}$$

Next, cell units (c) will be updated into cell state as seen in Equation (5).

$$\vec{c}_t = \vec{f}_t \circ \vec{c}_{t-1} + \vec{i}_t \circ \tanh\left(\vec{x}_t W_{xc} + \vec{h}_{t-1} W_{hc} + \vec{b}_c\right) \tag{5}$$

where ($\circ$) denotes an element-wise multiplication.

The parts of the acquired cell states will be cut by output gate. Below is the equation of the output gate (o).

$$\vec{o}_t = \sigma\left(\vec{x}_t W_{xo} + \vec{h}_{t-1} W_{ho} + \vec{c}_t W_{co} + \vec{b}_o\right) \tag{6}$$

The output value from the cell, which is the hidden activation, is formed from the calculation of the two terms as seen in Equation (7).

$$\vec{h}_t = \vec{o}_t \circ \tanh\left(\vec{c}_t\right) \tag{7}$$

There were three LSTM models compared in this research, including Vanilla LSTM, Stack LSTM and Bidirectional LSTM. Generally, the structure of the three LSTM models could be seen in Figure 3. Vanilla LSTM was the basic LSTM configuration [16] by adding densely connected as an additional layer applied to reducing network complexity and improving accuracy. Stack LSTM is a deep learning model that is done by layering some hidden LSTM so that they form a deeper model that is capable of getting more accurate object descriptions [16].

Meanwhile, Bidirectional LSTM combined two ways LSTM, Forward and Backward, to improve the performance of the network. In Forward LSTM layer, the input was arranged just the way it was, while in Backward LSTM, the input was arranged in a descending order from the available data. In Bidirectional LSTM, element-wise sum was used to combine outputs from Forward LSTM and Backward LSTM as in Equation (8) below.

$$h_i = \left[\vec{h}_i \oplus \overleftarrow{h}_i\right] \tag{8}$$

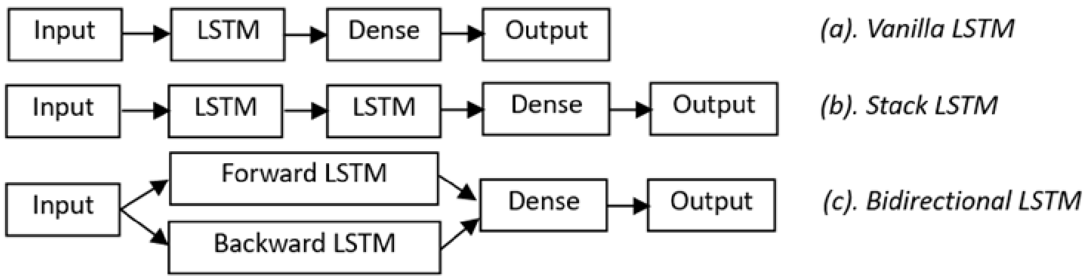

Figure 3. Three LSTM models that were compared

This approach gave relatively significant influence in LSTM recurrent neural network. It provided arrangements for both forward and backward, that were practically needed when the model used input vectors.

3. **Result and Discussion.** This study compared the proposed model (Bidirectional LSTM) with Vanilla LSTM and Stack LSTM. In the first step, Vanilla LSTM was implemented to create a prediction of rice crops pest attack. Next, sliding windows were applied to Vanilla LSTM to find the best size of sliding window. Table 1 below shows the comparison of Vanilla LSTM model performances by applying sliding windows of various sizes.

Table 1. The application of sliding windows in Vanilla LSTM

Model	**Window length**	**Training duration** (second)	**Training score**		**Testing score**	
			MSE (±)	RMSE (±)	MSE (±)	RMSE (±)
Vanilla LSTM	5	5.893	0.149	0.386	0.144	0.379
	10	5.499	0.143	0.378	0.138	0.372
	15	4.160	0.142	0.377	0.134	0.366
	20	4.148	0.137	0.370	0.136	0.369
	25	4.121	0.139	0.374	0.136	0.370

The best performance in this experiment could be seen from the indicator of training durations, training scores, and testing scores. Training and testing scores could be seen from MSE and RMSE values. The smaller the MSE and RMSE values, the better the system's performance. The searching of the optimum windows length value created model performance graphic, as presented in Figure 4.

Figure 4 shows that the best performance during training was obtained when the window length is 20. However, the best performance during testing was achieved at the window length 15. Based on the experiment result, we can infer that windows length has significant influence on the performance of the proposed model and its acceptable values are in the range of 15 to 20.

After obtaining the optimal length of windows, the next experiment was setting parameters such as epoch and batch size. After several experiments, it turns out that the best parameter setting was at epoch = 50 and batch size = 100. The results of these parameter settings appeared on the graph of the loss model, as can be seen in Figure 5. The graph shows that the value of training loss decreased continuously, indicating that the model was well trained.

The last step of this experiment was comparing Vanilla LSTM and Stack LSTM with Bidirectional LSTM. Table 2 shows the results of the comparisons. From the table, it can be seen that in terms of training duration, Stack LSTM and Bidirectional LSTM took longer time to perform training process. However, in terms of accuracy performance, it

FIGURE 4. Model performance by windows length

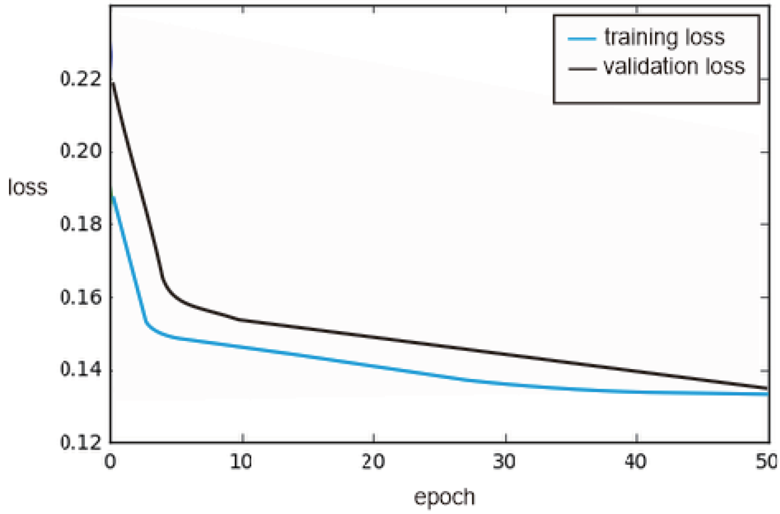

FIGURE 5. Model loss graphic

TABLE 2. The comparison among Vanilla LSTM, Stack LSTM and Bidirectional LSTM

Model	**Window length**	**Training duration** (second)	**Training score**		**Testing score**	
			MSE (±)	RMSE (±)	MSE (±)	RMSE (±)
Vanilla LSTM	20	4.148	0.137	0.370	0.136	0.369
Stack LSTM	20	6.226	0.137	0.371	0.137	0.370
Bidirectional LSTM	20	7.429	0.133	0.365	0.135	0.368

can be concluded that Bidirectional LSTM had the highest accuracy level compared to Stack LSTM and Vanilla LSTM. This can be seen in the MSE and RMSE values of the results in the training and testing process in those models.

4. **Conclusions.** The setting of the appropriate windows lengths in split windows methods gave a significant influence toward multivariate time series forecasting models that used LSTM. However, the optimum value needed to be considered since the wider the window did not always mean the better the performance. The use of Stack LSTM that was done by layering some hidden LSTM layers to make a deeper model could improve the accuracy value of the model, although it needed longer duration. The best accuracy

performance to predict rice crops pest attack using multivariable was gained by applying Bidirectional LSTM, which combined Forward and Backward LSTM so that the training process could be done thoroughly for both forward and backward input data arrangements. Further studies can be done by optimizing data preprocessing steps to be able to capture the characteristics of multivariate time series data. A sensitivity analysis is also suggested in order to understand the effectiveness of the proposed model. Comparison of models with other LSTM variants also needs to be done, for example with Gated Recurrent Units (GRU) which has two gates. Moreover, the number of the dataset can also be added to improve the performance of the model.

Acknowledgment. The authors would like to thank The Center for Food Crop and Horticulture Protection of Central Java Province and The Weather Observation Station of Indonesia Air Force in Adisumarmo Airport that have given the permission to access the dataset for this research.

REFERENCES

[1] S. Y. J. Prasetyo, S. Subanar, E. Winarko and B. S. Daryono, ESSA: Exponential smoothing and spatial autocorrelation methods for prediction of outbreaks pest in Indonesia, *International Review on Computers and Software (IRECOS)*, vol.10, no.2, pp.362-371, 2015.

[2] F. F. Rachman, 63000 hectares of rice fields attacked by planthopper pests, *Detik Finance Magazine*, 2017.

[3] Y. Zhuo, J. You, H. Xue and J. Wang, F-measure: A forecasting-led time series distance measure in large-scale forecasting of video services performance, *International Journal of Innovative Computing, Information and Control*, vol.14, no.6, pp.2175-2188, 2018.

[4] X. Wu, C. Wang, Z. Xu and Y. Hu, Application of GIS technology in monitoring and warning system for crop diseases and insect pests, *Geoinformatics and Joint Conference on GIS and Built Environment: Monitoring and Assessment of Natural Resources and Environments*, 2018.

[5] D. C. Hooker, A. W. Schaafsma and L. Tamburic-Ilincic, Using weather variables pre- and post-heading to predict deoxynivalenol content in winter wheat, *Plant Disease*, vol.86, pp.611-619, 2011.

[6] M. I. Indriyani, F. Nhita and D. Saepudin, Prediction of spread the drill stem pest based on weather information using adaptive neuro-fuzzy inference system algorithm in Kabupaten Bandung, *e-Proceeding of Engineering*, vol.3, no.2, 2016.

[7] A. K. Tripathy et al., Data mining and wireless sensor network for agriculture pest/disease predictions, *World Congress on Information and Communication Technologies*, pp.1229-1234, 2011.

[8] C. B. Pérez-Ariza et al., Prediction of coffee rust disease using Bayesian networks, *Proc. of the 6th European Workshop on Probabilistic Graphical Models*, pp.259-266, 2011.

[9] K. Klem et al., A neural network model for prediction of deoxynivalenol content in wheat grain based on weather data and preceding crop, *Plant Soil and Environment*, vol.53, pp.421-429, 2007.

[10] T. B. Wibowo and Sutikno, Paddy pest attacks prediction using back propagation neural networks, *Jurnal Teknik Informatika*, vol.9, no.2, pp.91-98, 2016.

[11] H. Goel, I. Melnyk and A. Banerjee, R2N2: Residual recurrent neural networks for multivariate time series forecasting, *arXiv.org*, arXiv:1709.03159, 2017.

[12] K. Y. Kim and C. Chung, On the evolution of the annual cycle in the tropical pacific, *Journal of Climate*, vol.14, no.5, pp.991-994, 2001.

[13] Q. Xiao, W. Li, P. Chen and B. Wang, Prediction of crop pests and diseases in cotton by long short-term memory network, in *Intelligent Computing Theories and Application. ICIC 2018. Lecture Notes in Computer Science*, D. S. Huang, K. H. Jo and X. L. Zhang (eds.), Cham, Springer, 2018.

[14] P. Maholtra, L. Vig, G. Shroff and P. Agarwal, Long short-term memory networks for anomaly detection in time series, *Proc. of the 23rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, Bruges, Belgium, 2015.

[15] M. S. AlDosari, *Unsupervised Anomaly Detection in Sequences Using LSTM-RNN*, Master Thesis, George Mason University, 2016.

[16] J. Brownlee, Long short-term memory networks with python: Develop sequence prediction models with deep learning, *Machine Learning Mastery*, 2017.

[17] N. Laptev, S. Smyl and S. Shanmugam, *Engineering Extreme Event Forecasting at Uber with Recurrent Neural Networks*, https://eng.uber.com/neural-networks/, 2019.

[18] M. F. Stollenga, *Advances in Humanoid Control and Perception*, Ph.D Thesis, Faculty of Informatics of the Università della Svizzera Italiana, 2016.