Published online on the journal's webpage: **http://jurnal.iaii.or.id**

Implementation of Convolutional Neural Network and Multilayer Perceptron in Predicting Air Temperature in Padang

Isman Kurniawan[1,2], Lusi Sofiana Silaban[1], Devi Munandar[3]
[1]School of Computing, Telkom University
[2]Research Center of Human Centric Engineering, Telkom University
[3]Research Center for Informatics, Indonesian Institute of Sciences
ismankrn@telkomuniversity.ac.id, lusisofianasilaban27@gmail.com, devi@informatika.lipi.go.id

Abstract

The air temperature is a physical parameter that affects many fields of daily life, such as agriculture, energy and medical. Hence, the ability to accurately predict the air temperature is necessary to support the operational processes in those fields. Regarding the necessity, this study aims to develop prediction models to predict the air temperature in Padang city, West Sumatera. The models were developed by using two variants of artificial neural network, i.e. Convolutional Neural Network (CNN) and Multilayer Perceptron (MLP) and the hybrid of those models. The data set used in this study is monthly air temperature from January 2015 to December 2017 measured at Lembaga Ilmu Pengetahuan Indonesia (LIPI) weather measurement station in Muaro Anai, Padang. The CNN model was developed by considering several parameters, such as filter number and kernel size. Meanwhile, several parameters considered in MLP are hidden layers number and neuron number. Those parameters were selected by using a hyperparameter tuning scheme. By using the optimized parameter, we found that the CNN model produce the most satisfying results with the value of R2 is 0.9965. This indicated that the CNN model is the best model to be used to predict air temperature.

Keywords: air temperature, CNN, MLP

1. Introduction

Air temperature is one of the weather parameters that have an important effect on daily life. Hence, the ability to predict the air temperature accurately is important in planning steps of a certain activity, such as flight recommendation, agriculture, sailing, etc. The prediction can be carried out by using two approaches, i.e. empirical approach and numerical approach. An empirical approach is performed by gathering data using observation of soil, satellite, etc. The data is forwarded to the meteorological center and then is converted to a multidimensional map by using a computer package. Meanwhile, a numerical approach utilizes the mathematical expression of weather variables to perform a prediction[1]. This approach is commonly performed by using several machine learning methods.

There are several studies performed by utilizing the machine learning (ML) method in predicting air temperature. An artificial neural network (ANN) is one of the ML methods that is commonly used to predict the air temperature. In 2009, Dombaycı and coworkers developed an ANN model to predict daily mean ambient temperatures in Denizli, Turkey, and found that ANN is reliable to be used for the prediction[2]. In 2015, Chithra and coworkers performed implemented ANN in a model to predict mean monthly maximum and minimum temperature in Kerala, India[3]. In 2015, Appelhans and coworkers predict monthly air temperature at Mt. Kilimanjaro, Tanzania by using 14 machine learning algorithms. By using 10-fold cross-validation, they found that regression trees produce better results than linear and non-linear regression models[4]. Other studies were also performed by implementing the ANN model to predict air temperature[5]–[7].

A prediction model to predict air temperature is also commonly developed by using a support vector machine (SVM). In 2011, Paniagua-Tineo and coworkers developed the SVM model to predict the daily maximum temperature in Europe, and the performance was compared to other methods[8]. In 2013, Mellit and coworkers predicted meteorological time series, including air temperature, by using least square SVM

and obtained promising results for short-term prediction[9]. Also, other studies have been performed by utilizing SVM in the prediction model[10], [11].

In this research, we performed air temperature prediction by using per day time-series data from January 2015 to December 2017. The data is taken from a measurement station of weather, owned by Lembaga Ilmu Pengetahuan Indonesia (LIPI), in Muaro Anai, Padang. We utilized three methods to predict the temperature, i.e. CNN, MLP and the hybrid of CNN-MLP. The hybrid method was developed by combining CNN and MLP model[12]–[14]. The performance of methods is evaluated by calculating RMSE and R^2 value.

2. Methodology

2.1. Data Set

We used a time-series temperature data that was taken from weather measurement station, owned by LIPI, in Muaro Anai, Padang, from January 2015 to December 2017. The location of the station is presented in Figure 1. The total number of the data is 1096, where the value of 130 data is missing. To treat this problem, we defined the data with missed value as the average value. We split the data into training and test data with a ratio of 67 % and 33 %, respectively.
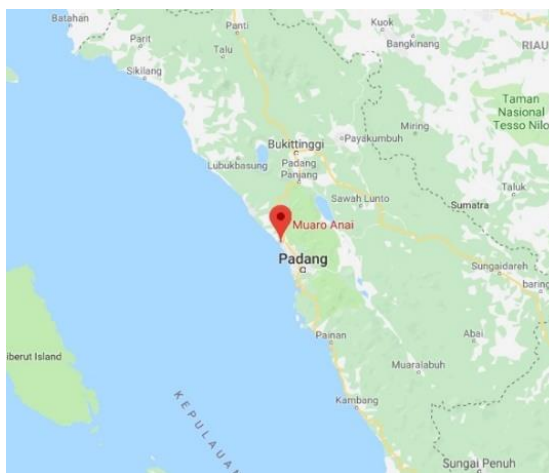

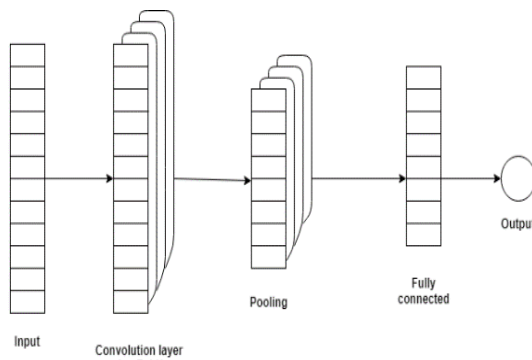Figure 1. The Location of Muaro Anai Weather Measurement Station

Figure 2. The Base Architecture of Convolutional Neural Network

2.2. Prediction Model

2.2.1 Convolutional Neural Network

Convolutional Neural Network is a variation of Artificial Neural Network (ANN). This method is originally proposed by Hubel and Wiesel in 1968 [15], which is usually used for image classification, semantic segmentation, object detection and feature extraction [16]. CNN consists of several layers where the structure is designed based on the case. The first layers involved in CNN are a convolutional layer comprised of neurons in the filter matrix form. The shifting of the filter matrix will be represented as a dot product or convolution process between input and kernel with a certain size, as shown in Figure 2. Then, the process will produce an output or called an activation map or feature map.

The second layer is the Rectified Linear Unit (ReLu) layer that converts a negative value to zero and maintains the positive value. The third layer is the Pooling layer that consists of a filter with a certain size and stride. In the pooling layer, the operation that is usually used is the maximum (max pooling) and average (average pooling) operation. The purpose of the usage of the pooling layer is to decrease the dimension of the feature map and as consequence increase the computation speed. In this research, we use max-pooling in the pooling layer. The last layer is a fully connected layer that reshapes feasture maps into the flatten input layer. This layer consist of the hidden layer, activation function, output layer, and loss function. This layer is also found in Multilayer Perceptron (MLP) that has an advantage in transforming the dimension to be classified in a linear manner[17]–[19].

2.2.2. Multilayer Perceptron

Multilayer Perceptron (MLP) is a kind of Artificial Neural Network (ANN) that uses a mathematical or computational model to process information based on the connectivity. The global behavior of this method is determined by the connection between the processing and parameters elements [20]. ANN is constructed to solve several problems, such as pattern recognition and classification. Generally, ANN has several kinds of structures, such as single-layer network, multilayer network and competitive layer network.

MLP consist of several neurons that are connected by synaptic weight. Those neurons are arranged in several layers, i.e. input layer, hidden layer, and output layer. MLP is a kind of feedforward neural network in which the information is move in one direction from the input layer, to the output layer via the hidden layer [21]. According to the structure, MLP can be classified as a multilayer network [22]. The MLP scheme is illustrated in Figure 3, where the input layer Y receives input from neuron x with synaptic weights w10...w14, w20...w24..., wn0...wn4. Then, the input will be sum up as formulated in Equation 1.
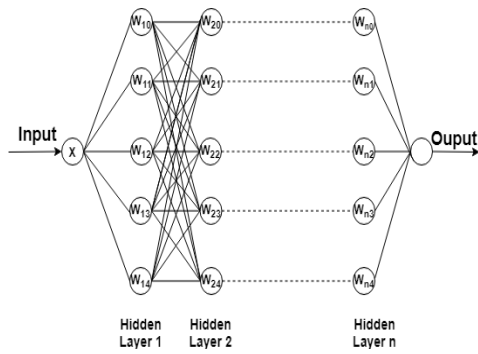
Figure 3. The Base Structure of Multilayer Perceptron

$$Y = xw_{10} + xw_{11} + \ldots + xw_{14} \tag{1}$$

The summation result of neuron and synaptic weight multiplication will be transformed by using the activation function that is used to determine the output in each neuron [23]. In this research, we consider using ReLU as an activation function. The ReLU activation function will give zero output if x < 0, or linear with gradient 1 for others. The output of the activation function will be neuron for the second hidden layer and so on [24]–[26].

2.2.3. Hyperparameter Tuning

Hyperparameter tuning is a process to find the best parameter of a model by defining a certain range of parameter values. In this research, we investigate the impact of the parameter on the RMSE value. Each combination of the parameters is evaluated by calculating the RMSE and R2 values. The hyperparameters of CNN and MLP that is tuned with the certain range values are provided in Table 1.

2.3. Evaluation

The performance of the prediction models of MLP and CNN with the best parameter was evaluated by calculating RMSE and R2 values, as expressed in Equation 2 and 3, respectively.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n}(x_i - y_i)^2} \tag{2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y_i})^2} \tag{3}$$

where x_i represents predicted value, y_i represents observed value, and n means the number of data [27].

3. Results and Discussions

3.1. Convolutional Neural Network

In the CNN method, we use three combinations of the convolutional layer (CL), i.e. 1, 2 and 4 layers, with a varied number of filters and kernel sizes. The combinations of kernel size and filter for each convolutional layer are (i) 24 filter - 8 kernel size (24f -

8ks) for 1 CL, (ii) 24f - 8ks, 12f - 4ks and 12f - 4ks for 2 CL, and 24f - 8ks, 12f - 4ks, 6f - 2ks and 3f - 1ks for 4CL. Those parameters were used in the training process to obtain a prediction model. From the combination of those parameters, we pick the best parameters and defined it as the best model. As shown in Figure 4, the fluctuation of RMSE value is different for each parameter.

Table 1. Hyperparameter of CNN and MLP Involved in Tuning Proccess

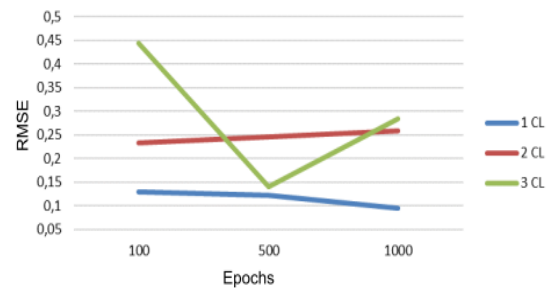| Method	Parameter	Range Vakues
Convolutional Neural Network	Epochs	[100, 500, 1000]
	Convolutional layer	[1, 2, 3, 4]
	Kernel size	[24, 12, 6, 3]
	Filter number	[8, 4, 2, 1]
Multilayer Perceptron	Hidden layer	[2, 3, 4]
	Epochs	[100, 500, 1000]
	Number of neurons	[5, 10]

Figure 4. The RMSE Values for the Combination of Convolutional Layer (CL) and Epochs in CNN Parameter Tuning

According to Figure 4, we found that CNN with 1 convolutional layer gives the best result with the RMSE value is 0.09. Figure 5 shows the plot of the predicted value against the actual value by using the CNN method. The plot is observed almost coincide which indicates that the model is well enough to recognize the time-series data.

These results can be related to the characteristic of CNN in extracting information. CNN has the ability to reduce the information by extracting a feature from part of the input. Hence, the more number of CNN means more information was reduced. In this case, the input size is not too large and thus 1 CNN layer is enough to extract necessary information. The addition of more CNN layer cause more information to dismiss.

3.2. Multilayer Perceptron

The prediction model with MLP is constructed by using the combination of the parameter as shown in Table 1. The best combination of the parameter was used to obtain the best model. The best model is determined by the comparison of RMSE for each model. Then, the best model was used to predict the test set as shown in Figure 6. Based on the figure, we found that the increase in epochs contributes to the increase in RMSE. However, in the case of the three hidden layers and ten hidden

neuron structures, the trend of RMSE is likely to be more fluctuated.

Based on Figure 6, we found that the lowest RMSE is obtained from MLP with two hidden layers and five hidden neuron structure. The plot of actual and predicted data by using the best model was provided in Figure 7. Based on the figure, we found that the plot of predicted data almost coincides against the actual data. This indicates that the model is also well enough to recognize the pattern of the time-series data.

These results can be related to the correlation between the number of hidden layers and hidden nodes with the model complexity. The more hidden layer and hidden node indicate the more complex of the model. In this case, the fewer number of hidden layers seems to produce the most accurate model. This is caused by the small architecture of ANN that can avoid overfitting conditions.
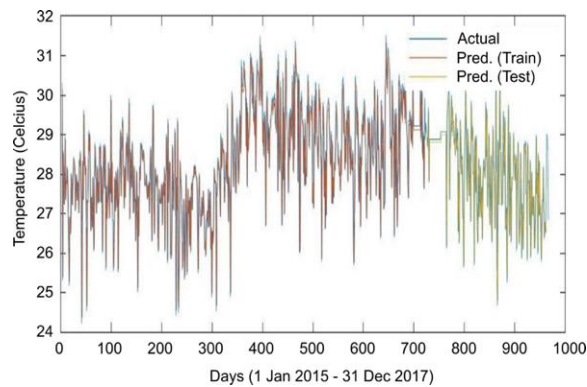

Figure 5. The Plot of Actual VS Predicted Values of Temperature by Using CNN
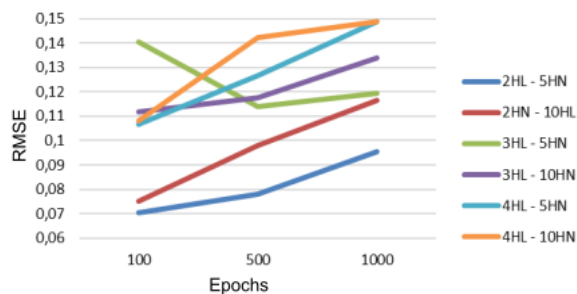

Figure 6. RMSE Values for the Combination of Hidden Layer (HL) and Hidden Nodes (HN) in MLP Parameter Tuning

3.3. The Hybrid of CNN-MLP

The values of temperature were also predicted by using the hybrid of the CNN-MLP method. In this method, we combine the implemented architecture of the best model of CNN and followed by the architecture of the best model of MLP. The architecture of MLP implemented after the flatten layer of CNN. The comparison of predicted and the actual value obtained from this model was presented in Figure 8. We found that the predicted line almost coincides with the actual line for both training and test data.
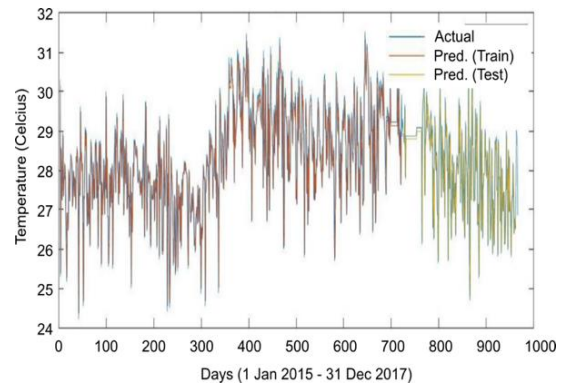

Figure 7. The Plot of Actual VS Predicted Values of Temperature by Using MLP Magnetization as a Function of Applied Field

3.4. The Comparison Model Performance

We compared the prediction accuracy of the best model of CNN, MLP and the hybrid of CNN-MLP. As for CNN, the best model is obtained from the architecture that involves one convolutional layer with 24 filters and eight kernel sizes. As for MLP, the best model is obtained from the architecture that involves two hidden layers and five hidden neurons. According to the result, we found that R2 of CNN, MLP and the hybrid of CNN-MLP are 0.9965, 0.9931, and 0.9941, respectively, as shown in Figure 9. The result shows that CNN presents the best performance in predicting the time series data of temperature used in this study.

The best performance of CNN is related to the ability of the method to extract the information of data in several kinds of patterns. Also, this method extracts the information by capturing a part of the data to reduce the dimension of data. However, the addition of MLP's architecture to CNN seems to decrease the performance of CNN, as found in the hybrid of CNN-MLP. This is related to the increase of the complexity of the model that also contributes to the decreasing of predictive ability on the test set.
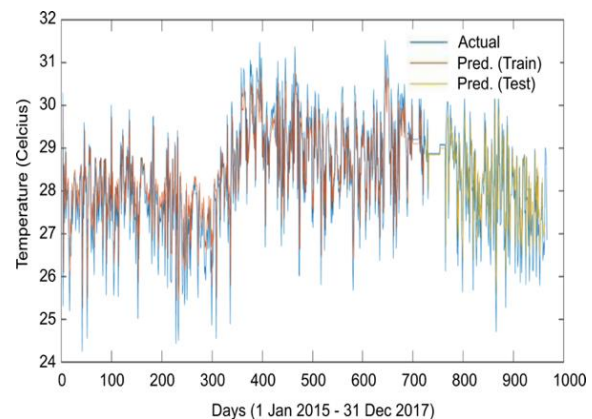

Figure 8. The Plot of Actual VS Predicted Values of Temperature by Using Hybrid of CNN-MLP

We also compared our results with reference [28] that utilized CNN and MLP in predicting air temperature. They also found that CNN produces the best

performance, even though the method is not commonly used in predicting air temperature with numerical input. Therefore, our results seem in agreement with the reference. This indicated that CNN can be used in predicting time series data of air temperature.

3.5 Limitations

There are limitations in our study that are related to the data set and model architecture. Our study used the data set of time series data of air temperature obtained from measurement station in Muaro Anai, Padang. Hence, the findings related to the methods might be different for other data set. Also, the hybrid method of CNN-MLP is developed by using the best architecture for each method. Other combinations of CNN and MLP architecture used in the hybrid method might produce different results.

4. Conclusion

In this research, we predict the time-series data of temperature by using CNN, MLP and the hybrid of CNN-MLP methods. The best model of CNN and MLP method were obtained by performing the hyperparameter tuning procedure. By using the best model, we predict both training set and test set and evaluate the performance by calculating RMSE and R2 value. From the results, we found that R2 of CNN is higher than that of MLP and CNN-MLP hybrid. This indicates that CNN presents the best performance in predicting the time-series data of temperature used in this study.
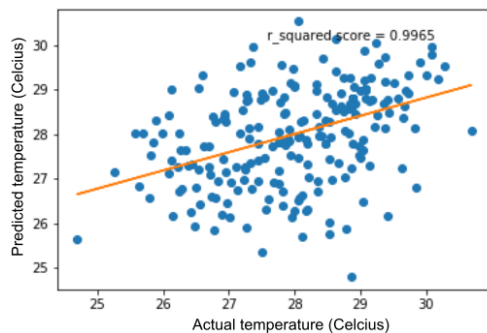
(a)
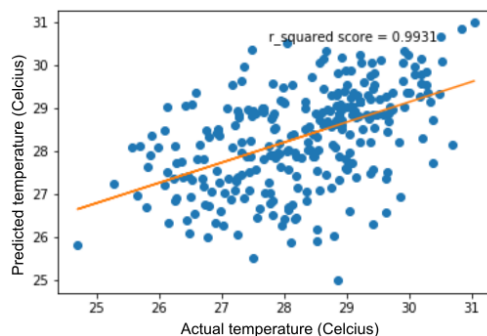
(b)
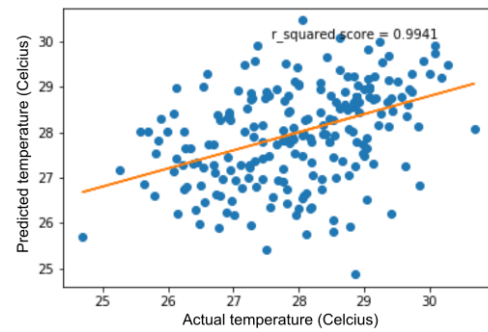
(c)

Figure 9. The Plot of Actual vs Predicted Values Obtained from (a) CNN, (b) MLP, and (c) Hybrid of CNN-MLP

References

[1] S. Khan, M. Muqeem, and N. Javed, "A Critical Review of Data mining Techniques in Weather Forecasting," *IJARCCE*, vol. 5, 2016.

[2] Ö. Altan Dombaycı and M. Gölcü, "Daily means ambient temperature prediction using artificial neural network method: A case study of Turkey," *Renewable Energy*, vol. 34, no. 4, pp. 1158–1161, Apr. 2009, doi: 10.1016/j.renene.2008.07.007.

[3] N. R. Chithra, S. G. Thampi, S. Surapaneni, R. Nannapaneni, A. A. K. Reddy, and J. D. Kumar, "Prediction of the likely impact of climate change on monthly mean maximum and minimum temperature in the Chaliyar river basin, India, using ANN-based models," *Theor Appl Climatol*, vol. 121, no. 3, pp. 581–590, Aug. 2015, doi: 10.1007/s00704-014-1257-1.

[4] T. Appelhans, E. Mwangomo, D. R. Hardy, A. Hemp, and T. Nauss, "Evaluating machine learning approaches for the interpolation of monthly air temperature at Mt. Kilimanjaro, Tanzania," *Spatial Statistics*, vol. 14, pp. 91–113, Nov. 2015, doi: 10.1016/j.spasta.2015.05.008.

[5] I. Tasadduq, S. Rehman, and K. Bubshait, "Application of neural networks for the prediction of hourly mean surface temperatures in Saudi Arabia," *Renewable Energy*, vol. 25, no. 4, pp. 545–554, Apr. 2002, doi: 10.1016/S0960-1481(01)00082-9.

[6] B. A. Smith, G. Hoogenboom, and R. W. McClendon, "Artificial neural networks for automated year-round temperature prediction," *Computers and Electronics in Agriculture*, vol. 68, no. 1, pp. 52–61, Aug. 2009, doi: 10.1016/j.compag.2009.04.003.

[7] B. Ustaoglu, H. K. Cigizoglu, and M. Karaca, "Forecast of daily mean, maximum and minimum temperature time series by three artificial neural network methods," *Meteorological Applications*, vol. 15, no. 4, pp. 431–445, 2008, doi: 10.1002/met.83.

[8] A. Paniagua-Tineo, S. Salcedo-Sanz, C. Casanova-Mateo, E. G. Ortiz-García, M. A. Cony, and E. Hernández-Martín, "Prediction of daily maximum temperature using a support vector regression algorithm," *Renewable Energy*, vol. 36, no. 11, pp. 3054–3060, Nov. 2011, doi: 10.1016/j.renene.2011.03.030.

[9] A. Mellit, A. M. Pavan, and M. Benghanem, "Least squares support vector machine for short-term prediction of meteorological time series," *Theor Appl Climatol*, vol. 111, no. 1, pp. 297–307, Jan. 2013, doi: 10.1007/s00704-012-0661-7.

[10] E. G. Ortiz-García, S. Salcedo-Sanz, C. Casanova-Mateo, A. Paniagua-Tineo, and J. A. Portilla-Figueras, "Accurate local very short-term temperature prediction based on synoptic situation Support Vector Regression banks," *Atmospheric Research*, vol. 107, pp. 1–8, Apr. 2012, doi: 10.1016/j.atmosres.2011.10.013.

[11] R. F. Chevalier, G. Hoogenboom, R. W. McClendon, and J. A. Paz, "Support vector regression with reduced training sets for

air temperature prediction: a comparison with artificial neural networks," *Neural Comput & Applic*, vol. 20, no. 1, pp. 151–159, Feb. 2011, doi: 10.1007/s00521-010-0363-y.

[12] F. Sharifzadeh, G. Akbarizadeh, and Y. Seifi Kavian, "Ship Classification in SAR Images Using a New Hybrid CNN–MLP Classifier," *J Indian Soc Remote Sens*, vol. 47, no. 4, pp. 551–562, Apr. 2019, doi: 10.1007/s12524-018-0891-y.

[13] C. Zhang *et al.*, "A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 140, pp. 133–144, Jun. 2018, doi: 10.1016/j.isprsjprs.2017.07.014.

[14] Y. Chu, C. Huang, X. Xie, B. Tan, S. Kamal, and X. Xiong, "Multilayer Hybrid Deep-Learning Method for Waste Classification and Recycling," *Computational Intelligence and Neuroscience*, Nov. 01, 2018. https://www.hindawi.com/journals/cin/2018/5060857/ (accessed Oct. 09, 2020).

[15] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex," *J. Physiol. Paris*, vol. 195, pp. 215–243, 1968.

[16] J. Wu, "Introduction to convolutional neural networks," National Key Lab for Novel Software Technology, Nanjing University, China: Semantic Scholar, 2017, pp. 5–23.

[17] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 International Conference on Engineering and Technology (ICET)*, Aug. 2017, pp. 1–6, doi: 10.1109/ICEngTechnol.2017.8308186.

[18] T. Guo, J. Dong, H. Li, and Y. Gao, "Simple convolutional neural network on image classification," in *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)*, Mar. 2017, pp. 721–724, doi: 10.1109/ICBDA.2017.8078730.

[19] T. Perol, M. Gharbi, and M. Denolle, "Convolutional neural network for earthquake detection and location," *Science Advances*, vol. 4, no. 2, p. e1700578, Feb. 2018, doi: 10.1126/sciadv.1700578.

[20] K. Gurney, *An introduction to neural networks*. CRC press, 2014.

[21] M. Riedmiller, "Advanced supervised learning in multi-layer perceptrons — From backpropagation to adaptive learning algorithms," *Comput. Stand. Inter.*, vol. 16, pp. 265–278, 1994.

[22] S. Haykin, *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1994.

[23] A. Marchisio, M. A. Hanif, S. Rehman, M. Martina, and M. Shafique, "A methodology for automatic selection of activation functions to design hybrid deep neural networks," 2018.

[24] R. F. MELLO and M. A. Ponti, *Machine Learning: A Practical Approach on the Statistical Learning Theory*. Springer, 2018.

[25] H. Ramchoun, Y. Ghanou, M. Ettaouil, and M. A. J. Idrissi, "Multilayer Perceptron: Architecture Optimization and Training," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, no. Special Issue on Artificial Intelligence Underpinning, 2016, Accessed: Oct. 09, 2020. [Online]. Available: https://www.ijimai.org/journal/bibcite/reference/2523.

[26] H. Ramchoun, M. A. J. Idrissi, Y. Ghanou, and M. Ettaouil, "Multilayer Perceptron: Architecture Optimization and training with mixed activation functions," in *Proceedings of the 2nd international Conference on Big Data, Cloud and Applications*, New York, NY, USA, Mar. 2017, pp. 1–6, doi: 10.1145/3090354.3090427.

[27] T. Chai and R. R. Draxler, "Root mean square error (rmse) or mean absolute error (mae)?–arguments against avoiding rmse in the literature," *Geosci. Model Dev.*, vol. 7, pp. 1247–1250, 2014.

[28] S. Lee, Y.-S. Lee, and Y. Son, "Forecasting Daily Temperatures with Different Time Interval Data Using Deep Neural Networks," *Applied Sciences*, vol. 10, no. 5, Art. no. 5, Jan. 2020, doi: 10.3390/app10051609.