E3S Web of Conferences **353**, 01006 (2022) https://doi.org/10.1051/e3sconf/202235301006
EVF'2021

Extreme Learning Machine versus Multilayer perceptron for rainfall estimation from MSG Data

Mourad Lazri[1*], *Fethi* Ouallouche[1], *Karim* Labadi[2], and *Soltane* Ameur[1]

[1] Laboratoire LAMPA (laboratoire d'analyse et de modélisation des phénomènes aléatoires) Faculty G.E.I, Mouloud MAMMERI University of Tizi-Ouzou (Algeria).
[2] ECAM-EPMI LR2E, / Quartz-Lab (EA7393), Cergy-Pontoise (France).

Abstract. The application of artificial neural networks (ANN) in several fields has shown considerable success for classification or regression. Learning algorithms such as artificial neural networks must constantly readjust during the learning phase. This requires a relatively long learning time compared to the size and dimension of the data used. Contrary to these considerations, a new neural network, such as Extreme Learning Machine (ELM) has recently been implemented. The ELM does not care much about the size of the neural network, the hidden layer parameters are randomly generated and remain constant instead of being adjusted during training. In this paper, we will present a comparison between two neural networks, namely ELM and MLP (Multilayer perceptron) implemented for the precipitation estimation from meteorological satellite data. The architecture chosen for the two neural networks consists of an input layer (7 neurons), a hidden layer (8 neurons) and an output layer (7 neurons). The MLP has undergone standard training as soon as the ELM is trained according to the characteristics mentioned above. The results show that MLP prevails over ELM. However, the time cost during learning is too high for MLP compared to ELM.

1 Introduction

Precipitation mapping is of great importance to farmers and hydrologists. Knowing the amounts of precipitation in time and space is valuable information for planning the management of water resources, especially in recent years when these resources are beginning to become scarce. However, the measurement of precipitation is a complex operation because, on the one hand, of the very great climatic variability accentuated by global warming, and on the other hand, of the insufficiency of spatial coverage by traditional measuring instruments based in ground, such as rain gauges and weather radar.
Observations from space are massively used to remedy the lack of data on the ground. They can provide complete and regular coverage and continuous monitoring of precipitation clouds. Since the advent of weather satellites, many techniques have been developed to

* Corresponding author: m_lazri@yahoo.fr

attempt to link satellite information and precipitation rates [1-3]. due to the indirect character of these data, the results strongly depend on the methods used. Machine Learning (ML) algorithms have shown their effectiveness in this area. One of the most powerful models in Machine Learning is the artificial neural network (ANN), in particular the multi-layer perceptron (MLP) which is very responsive and widely used [4].

However, ANNs in general require iterative tuning of network parameters, especially those of the hidden layers. This leads to a relatively long learning due to iterative calculations of gradient to adjust the parameters. Despite the level of precision provided by these neural networks using the gradient during learning, sometimes they do not always give the best overall solution. They depend on the complexity of the feature space and especially on the initialization of the network parameters. Therefore, the approximation solution can converge at local extrema. In contrast to these gradient-based techniques, the Extreme Learning Machine Neural Network proposed by Guang-Bin and Qin-Yu [5], aims to train single-layer hidden feedforward networks (SLFNs). It consists of randomly assigning coefficients to the weights between the input layer and the hidden layer and the biases in the hidden layer with nonlinear activation functions in the hidden layer. No adjustments are made during learning. The only parameter to learn for the network is the coefficient assigned to the weight between the hidden layer and the output layer.

Thus, the ELM network which learns without iteration converges much faster unlike traditional neural networks. According to some works, this characteristic allows ELM to reach a global optimal solution [6]. The ELM benefits from a superior learning speed and a good ability to generalize [7] hence its use in a variety of learning problems, such as classification, regression…

The ELM that overcomes the slow training speed and the over-adjustment problem, however, may cause the classification performance to fluctuate. Based on these two points, as part of this work, we will make a comparison between the two neural networks ELM and MLP in the classification and estimation of precipitation from MSG (Meteosat Second Generation) data. Training and validation of ELM and MLP is performed by comparing MSG data to radar data.

The rest of the paper is organized as follows. Section 2 presents the study area and the data used. Section 3 briefly describes the mathematical concept of ELM and MLP. The classification and estimation results are presented in section 3. The conclusion and outlook are given in section 4.

2 Study area and data

As part of this work, rainfall was estimated over northern Algeria (see Fig.1) using data from the MSG satellite.

We used MSG data for the classification and estimation of precipitation and radar data for learning and validation of the implemented model. MSG data are available as multi-spectral images (12 channels) at a frequency of 15 minutes with a spatial resolution of 3km for 11 channels and 1km for a high resolution visible channel [8].

As for the radar data, a radar image is provided every 15 minutes whose pixels can take 16 levels of precipitation intensities in dBZ (04, 12, 18, 22, 26, 30, 34, 38, 42, 46, 50 , 54, 58, 62, 66, 70). The spatial resolution is 1 km. For the spatial correspondence of the two types of data, we transformed the radar resolution from 1x1 km2 to 4x5 km2 by taking the average of the pixels. The 4 x 5 km2 resolution corresponds to the satellite resolution in the study region.

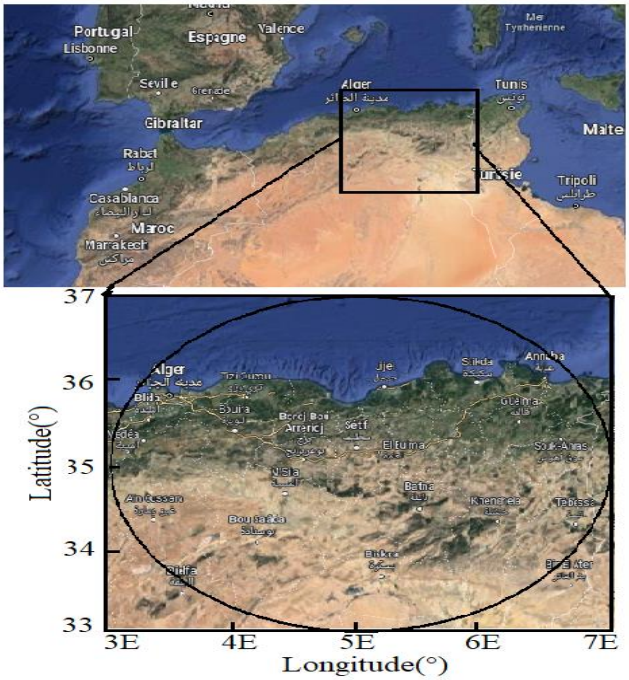

Fig. 1. Study area and radar coverage (circle).

To better establish the relationship between the satellite data and the precipitation intensities, we have chosen the combinations given in Table 1 providing the information on the optical and microphysical properties of the clouds from the different channels of MSG [9-11]. Target output classes are also given in Table 1.

Table 1. Input and output parameters for the two models (MLP and ELM).

Input	**Cloud Optical and Microphysical Properties**	**Output (dBZ)**	**Average rate (mm/h)**
$T_{10.8}$ (Kelvin)	CTH (Cloud Top High) CTT (Cloud Top Temperature)	62 ≤ Class(1) ≤ 70	09.32
$\Delta T_{10.8-12.0}$ (Kelvin)	Cloud Phase (CP)	54 ≤ Class(2) ≤ 58	07.14
$\Delta T_{8.7-10.8}$ (Kelvin)	Cloud Phase (CP)	42 ≤ Class(3) ≤50	06.41
$\Delta T_{6.2-10.8}$ (Kelvin)	CTH (Cloud Top High) CTT (Cloud Top Temperature)	34 ≤ Class(4) ≤ 38	03.60
$\Delta T_{7.3-12.1}$ (Kelvin)	CTH (Cloud Top High) CTT (Cloud Top Temperature)	22 ≤ Class(5) ≤ 30	02.02
$R_{0.6}$ (μm) Daytime	Cloud Water Path (CWP)	12 ≤ Class(6) ≤ 18	00.89
$R_{1.6}$ (μm) Daytime	Cloud Water Path (CWP)	Class(7) = 04	00.00
$\Delta T_{3.9-10.8}$ (Kelvin) Nighttime	Cloud Water Path (CWP)		
$\Delta T_{3.9-7.3}$ (Kelvin) Nighttime	Cloud Water Path (CWP)		

3 Mathematical concept of MLP and ELM

3.1 Multilayer perceptron (MLP)

The MLP type neural network consists of a structure of successive layers including an input layer and an output layer connected one after the other [12]. Between the input layer and the output layer, the MLP can have one or more hidden layers. A layer is a set of neurons that feed the neurons of the next layer through their outputs. To work, each input x_i of a neuron j is multiplied by an adaptive coefficient w_{ij} (synaptic weight). Then, the nonlinear activation function is applied at the level of the output neurons. This function usually takes the form of a sigmoid $\emptyset$. Thus, the output of the jth neuron, oj, is given by:

$$O_j = \emptyset\left(\sum_{i=0}^{n} w_{ij}.x_i\right) \quad (1)$$

where n is the number of variables in the dataset. The learning process consists of adjusting the synaptic weights of the MLP. In our case, we used the Levenberg-Marquardt and Gauss-Newton learning algorithms. This algorithm allows a stable and fast convergence [13].

3.2 Extreme Learning Machine (ELM)

The Extreme Learning Machine (ELM) is an extremely fast learning neural network [5], which can be an advantage for some applications. ELM is a feed-forward type algorithm whose architecture is identical to MLP with a single hidden layer. The main difference is the learning operation. Indeed, unlike MLP, the hidden layer parameters of ELM are randomly generated and remain constant instead of being readjusted.

The input layer is a passage of the input data inside the network where the number of neurons in this layer is in general the same as the size of the feature vector. The number of neurons in the output layer is equal to the number of output target classes in the classification case. As for the number of neurons in the hidden layer, it is variable. As part of this study, we will see its influence on the quality of the classification. Fig. 2 presents the general structure of an ELM network.

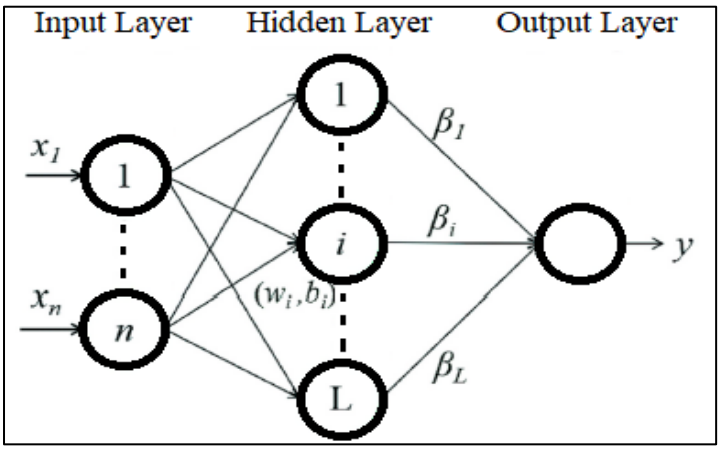

Fig. 2. General architecture of a neural network (ELM type)

The w_i represent the link weights between the input layer and the hidden layer. The β_i are the weights assigned to the link between the hidden layer and the output layer. The b_i are the biases of the hidden layer.

The Algorithm produces a random connection weight between *N* arbitrary input samples and the *L* hidden layer nodes. To activate this function, it is necessary to give the number of hidden neurons. Mathematically, the ELM model can be expressed according to the equation 2:

$$\sum_{i=1}^{M} \beta_i g(x_n; b_i, w_i) = y_n \; with \; n = 1,2, \dots, N \tag{2}$$

The $g(x_n; b_i, w_i)$ is the output result of the input sample x_n related to the *ith* hidden node. All inputs are randomly distributed to nodes in the hidden layer of the ELM. Which gives the following expression:

$$H\beta = Y \tag{3}$$

Where

$$H = \begin{bmatrix} g(x_1; b_1, w_1) & \dots & g(x_1; b_M, w_M) \\ & \vdots & \\ g(x_N; b_1, w_1) & \dots & g(x_N; b_M, w_M) \end{bmatrix} \tag{4}$$

The H is the matrix of size N*M

$$H\beta = (\beta_1^T, \beta_2^T, \dots, \beta_L^T)_{m \times M} \tag{5}$$

And the output (Y) is given by

$$Y = (t_1^T, t_2^T, \dots, t_L^T)_{m \times M}^T \tag{6}$$

According to *Zhen et al*. [14], "*H*" is the hidden layer output matrix and "*T*" is the label matrix.

The β is the output weights, which are obtained by determining the least-squares solutions to the linear system described above [15]:

$$\beta = H^+ Y \tag{7}$$

Deo and Sahin [15] define "*H*+" as the generalized inverse of the Moore-Penrose *H* matrix.

For the learning of ELM, we carried out the following steps:

- *Training set :*

$$S = [(x_i, y_i) | \; x_i \in R^n, y_i \in R^m, i = 1, \dots, N] \tag{8}$$

- *Initialization :*

Random assignment of values to the hidden layer weights w_i and bias b_i and calculation of the hidden layer output matrix *H* from training set.

- *Analytical solution*

Obtain β from $H\beta=T$ by Moore Penrose inverse. The $\beta=H^+T$, where H^+ is the Moore-Penrose generalized inverse of matrix *H*.

For the classification and estimation of precipitation, we applied the two types of neural networks, namely MLP and ELM. The inputs of both models are MSG data (see Table 1) and the target outputs are classes of precipitation intensities. Both models are trained by comparing the MSG input data to the Radar output data for the rainy season period from October 2007 to March 2008. We varied the number of neurons in the hidden layer for the ELM and the number of layers hidden for MLP.

4 Results

4.1 Precipitation classification

In this section, we will give classification results and precipitation estimates by applying ELM and MLP neural networks. Thus, the precipitation scenes collected during the period from October 2008 to March 2009 are classified. The classification accuracy is calculated by comparing the classification results with the reference data from the radar. For both models, we used the same architecture, namely 7 nodes for the input layer, 8 nodes for the hidden layer and 7 nodes for the output layer. We calculated the classification accuracy for the two models for the different classes (see Fig. 3).

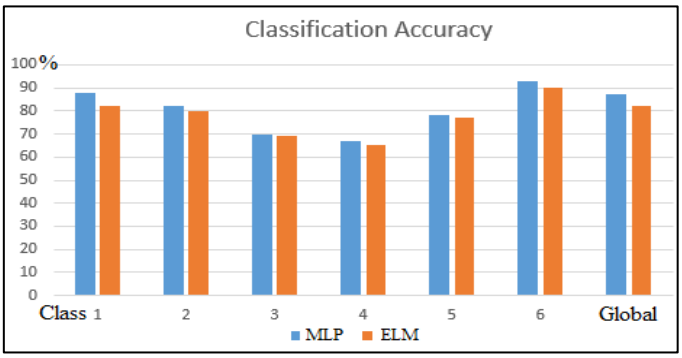

Fig. 3. Classification accuracy for MLP and ELM

As shown in Figure 3, the MLP neural network gives the best good classification rates for all classes. For visual illustration, we have classified an instantaneous precipitation scene taken on January 24, 2009 at 3 p.m. (see Fig. 4).

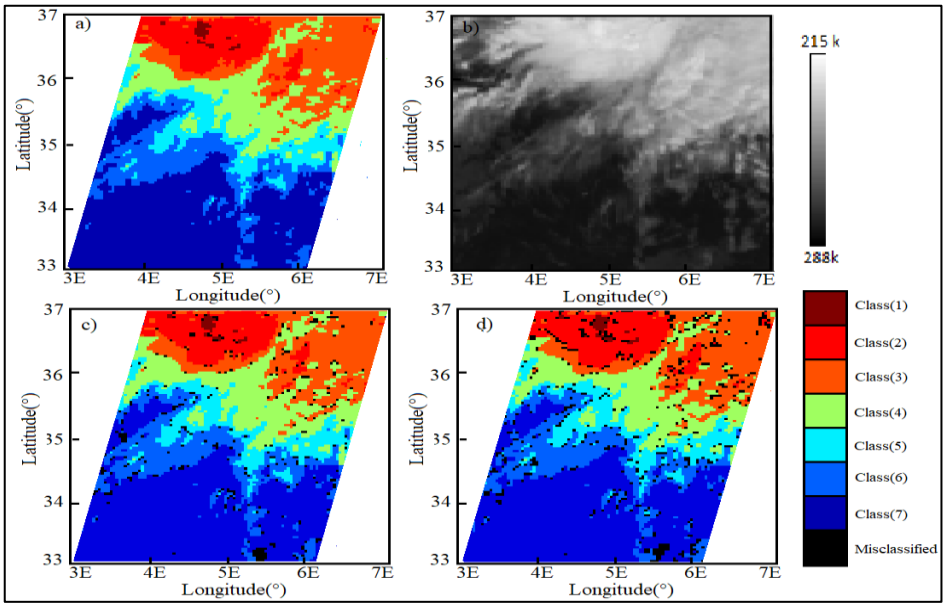

Fig. 4. Classification results, a) Reference image from radar, b) Original IR10.8 image, c) Classified image using MLP, d) Classified image using ELM

The number of misclassified pixels is very present in the classification results obtained by applying ELM compared to the result obtained by MLP.

To see the influence of the number of hidden layers on the results for MLP and the influence of the number of hidden neurons for ELM, we varied these parameters. The results are shown in Fig. 5.

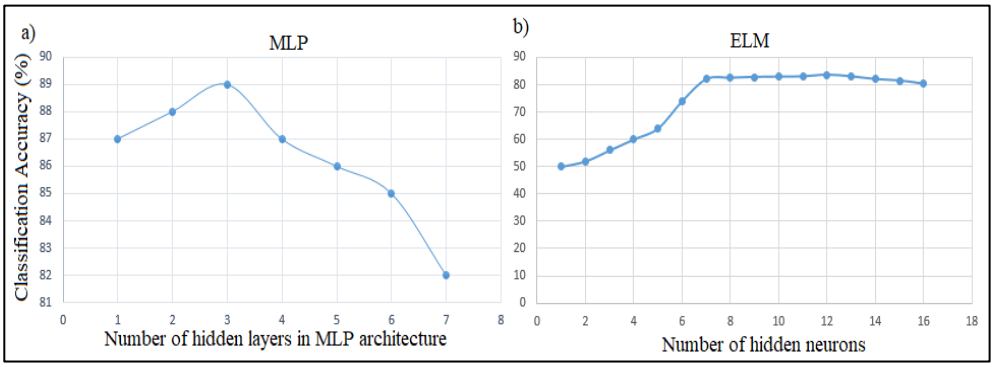

Fig. 5. Classification accuracy, a) depending to number of hidden layers performed by MLP, b) depending to number of hidden neurons performed by ELM

As shown in Fig. 5, for the MLP (see Fig.5.a), the increase in the number of hidden layers has practically no effect on the classification, even from the third hidden layer, the rate of good classification has experienced a decrease. In the case of ELM (see fig.5.b), the number of hidden neurons can influence the results. We noted an improvement in the quality of classification. However, from 7 to 8 hidden neurons, the results deteriorate.

4.2 Precipitation estimates

In this part, we present the results of precipitation estimates from MLP and ELM classifications using the equation:

$$RR(mm) = \frac{1}{4}\sum_{i=1}^{7} Nclass(i) \times R(i) \tag{9}$$

RR(mm) is the precipitation rate for a given period. *Nclass(i)* is the number of occurrences of *class(i)* during the period and *R(i) (mm/h)* is the average precipitation intensity of *class(i)*.

In order to compare the estimation results between MLP and ELM, the same architecture was used for both models. The results in the form of a fluctuation diagram are given in fig.6. Evaluation parameters, such as Bias, root mean square error (RMSE) and correlation coefficient (CC) are calculated using the equations (10), (11) and (12) (see Table 2)

$$\text{Bias} = \frac{1}{\text{N}}\sum_{\text{i=1}}^{\text{N}}(\text{E}_\text{i} - \text{V}_\text{i}) \tag{10}$$

$$\text{RMSE} = \sqrt{\frac{1}{\text{N}}\sum_{\text{i=1}}^{\text{N}}(\text{E}_\text{i} - \text{V}_\text{i})^2} \tag{11}$$

$$\text{CC} = \frac{\sum[(E_i - \bar{E}) * (V_i - \bar{V})]}{\sqrt{\sum(E_i - \bar{E})^2 * \sum(V_i - \bar{V})^2}} \tag{12}$$

Where E_i is the estimated value, and V_i is the actual values obtained from data Radar.

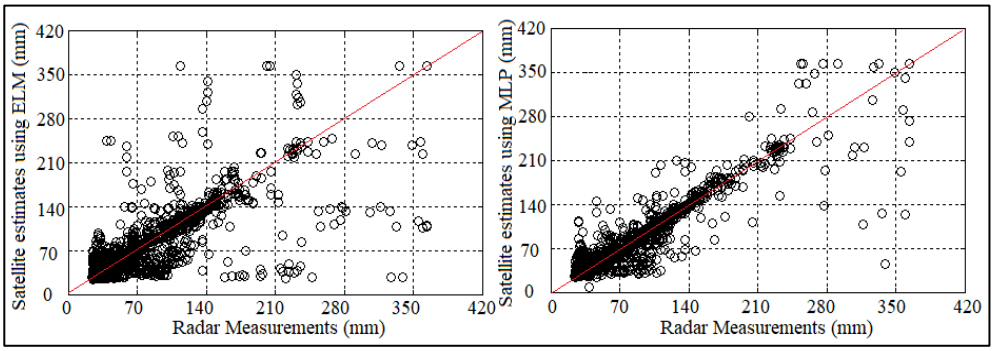

Fig. 6. Rain estimation versus radar measurements, (left) using ELM, (right) using MLP.

Table 2. Statistical scores

	CC	RMSE (mm)	Bias (mm)
MLP	0.81	24	-5.1
ELM	0.73	31	-7.2
Optimal	1	0	0

The results clearly show that:

- The MLP wins over ELM. Indeed, the estimates by MLP are well correlated with the radar measurements for the MLP compared to ELM. The CC shows 0.81 and 0.73 for MLP and ELM respectively. In RMSE terms, the error is larger in the case of ELM (31mm) versus MLP (24mm). Both models show an underestimation of precipitation. It is -7.2mm for ELM against -5.1mm for MLP.
- Nevertheless, MLP has a disadvantage regarding its a high computational cost during learning. We noted the average time required for learning 33254 to 40241 seconds for the MLP against 4120 to 7542 seconds for the ELM. A compromise can be found between the quality of the results and the learning time, depending on the type of applications.

5 Conclusion

The purpose of this work is, on the one hand, to estimate rainfall from satellite data in northern Algeria, and on the other hand to test the contribution of ELM compared to MLP. Both models were therefore implemented and used for precipitation classification and estimation. Each of the architectures of the two models consists of an input layer, a hidden layer and an output layer. For the implementation of this study, multi-spectral images from MSG satellites and their correspondences in radar images are combined and used. MSG data provide information on the optical and microphysical properties of clouds. From these data, seven input parameters are obtained and at the output seven target classes are selected.

The classification results showed the superiority for the MLP against the ELM. For the different classes, the classification accuracy for MLP exceeds that of ELM. In the case of the precipitation estimate, the same trend was observed. The different evaluation parameters, such as CC, Bias and RMSE show that MLP is better than ELM. However, the learning time for the MLP is a major drawback, especially for certain applications operating in real time and their readjustment must be done regularly. ELM may be an adequate solution in the near future as many variants of ELM are proposed to further improve its stability and generalization for specific applications.

References

1. D.W. Goshime, R. Absi, B. Ledésert, Evaluation and Bias Correction of CHIRP Rainfall Estimate for Rainfall-Runoff Simulation over Lake Ziway Watershed, Ethiopia, Hydrology, **6(3)**, 68.
2. D.W. Goshime, R. Absi, A.T. Haile; B. Ledésert, T. Rientjes 2020 "Bias-Corrected CHIRP Satellite Rainfall for Water Level Simulation, Lake Ziway, Ethiopia", Journal of Hydrologic Engineering, Volume **25** Issue 9.
3. M. Lazri, K. Labadi, J.M. Brucker, S. Ameur, Improving satellite rainfall estimation from MSG data in Northern Algeria by using a multi-classifier model based on machine learning, Journal of Hydrology, Volume **584**, 2020, 124705,
4. K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," Neural networks, vol. **2**, no. 5, pp. 359–366, 1989. 8.
5. Guang-Bin H, Qin-Yu Z, Chee-Kheong S (2006) Extreme learning machine: Theory and applications. Neurocomputing **70**(1-3):489–501.
6. Leung HC, Leung CS, Wong EWM (2019) Fault and Noise Tolerance in the Incremental Extreme Learning Machine. IEEE Access 7:155171–155183.
7. Li H-T, Chou C-Y, Chen Y-T, Wang S-H, Wu A-Y (2019) Robust and Lightweight Ensemble Extreme Learning Machine Engine Based on Eigenspace Domain for Compressed Learning. IEEE Trans Circuits Syst I: Regular Papers **66**(12):4699–4712.
8. EUMETSAT, 2004. Applications of Meteosat Second Generation – Conversion from Counts to Radiances and from Radiances to Brightness Temperatures and Reflectance, http://oiswww.eumetsat. org/WEBOPS/msg_interpretation/index.html.
9. Thies, B., Nauss, T., Bendix, J., 2008. Precipitation process and rainfall intensity differentiation using Meteosat second generation spinning enhanced visible and infrared imager data. J. Geophys. Res. **113**.
10. Lazri, M., Ouallouche, F., Ameur, S., Brucker, J.M., Mohia, Y., 2012. Identifying convective and stratiform rain by confronting SEVERI sensor multispectral infrared to radar sensor data using neural network. Sens. Transducers J. **145** (10), 19–32.
11. Feidas, H., Giannakos, A., 2011. Identifying precipitating clouds in Greece using multispectral infrared Meteosat Second Generation satellite data. Theor. Appl. Climatol. **104**, 25–42.
12. Haykin, S.: Neural Networks and Learning Machines, 3rd edn. Prentice-Hall (2009)
13. Yu, H., Wilamowski, B.: The Industrial Electronics Handbook, vol. **5**. CRC (2011).
14. Zhen Nan, L., Li, Q.F., Nguyen, L.B., Xu, G.H., 2018. Comparing machine-learning models for drought forecasting in vietnam's cai river basin. Pol. J. Environ. Stud. **27** (6), 2633–2646.
15. Deo, R.C., S¸ahin, M., 2015. Application of the extreme learning machine algorithm for the prediction of monthly Effective Drought Index in eastern Australia. Atmos. Res. 153, 512–525. https://doi.org/10.1016/j.atmosres.2014.10.016, 2015.