Prediction of Forest Fire using Neural Network based on Extreme Learning Machines (ELM)

Mochammad Anshori
Faculty of Computer Science
Brawijaya University
Malang, Indonesia
aanshori@student.ub.ac.id

Farhanna Mar'i
Faculty of Computer Science
Brawijaya University
Malang, Indonesia
farhannamar@student.ub.ac.id

Mukhammad Wildan Alauddin
Faculty of Computer Science
Brawijaya University
Malang, Indonesia
mwildanalauddin@student.ub.ac.id

Wayan Firdaus Mahmudy
Faculty of Computer Science
Brawijaya University
Malang, Indonesia
wayanfm@ub.ac.id

***Abstract*— To prevent forest fires, predictions need to be made to find out areas of land that have the potential to burn based on meteorological conditions obtained from the sensor, so that it is expected to reduce the spread of fire before the fire spreads. Meteorological conditions used in this study to predict areas of land that will be affected by forest fires are temperature, wind, humidity, and rainfall. The method used in this study is a neural network with Extreme Learning Machines (ELM) training model. To improve the performance of the ELM method, in this study several tests will be carried out so that the resulting predictions are the best.**

Keywords— prediction, forest fire, extreme learning machine

I. INTRODUCTION

Forest fires are natural disasters which are important problems that have become of international concern . As a source of oxygen, forests have an important role in natural balance. Fire events that occur in forests can have a significant negative impact on the environment, including the killing of living things in the forest, destruction of ecosystems, and smog resulting from fires that are very detrimental to humans. Carbon emissions that occur due to forest fires will endanger humans and also other living creatures that are around the forest [1]. Some of the major fire events that have occurred in the world include a wildfire that occurred in China counted from 1952 to 2012, 798,500 forest fires have occurred and caused 38,060,000 hectares of forest land to be damaged and losses of $ 22 million a year and 13,000 forest fires in every year [2]. Whereas in Portugal almost 20,000 forest fires occur every year, in Russia every year 7 million hectares of land are burned, and in Indonesia since 1990 there have been 27.5 million hectares of land burned [3].

Strategies to deal with forest fires can be carried out with efforts to prevent, detect, and also predict forest fires. Prediction efforts are carried out to find out in advance areas that are potentially and at greater risk of being affected by forest fires so that they can enable preventive actions by knowing the spread of wildfire that will occur [4]. To make predictions, it is necessary to know first the factors that cause forest fires. Forest fires can occur due to several factors, one of which will be used in this study is natural factors. These natural factors include meteorological conditions such as temperature, wind, humidity, and rainfall which will affect the occurrence of forest fires as well as wildfire indices or FWI (Fire Weather Index)[5]. The parameter's value of the meteorological condition can be used to predict the number of areas of land that will burn [6], so that with the prediction it is expected that prevention can be made of areas prone to fire and also minimize the spread of wildfire. In this study forest prediction will be carried out based on meteorological conditions parameters, namely temperature, wind, humidity and rainfall by output of the number of areas affected by the fire. The data to be used is a report of forest fires in Montenshino Natural Park in Portugal, and there are 517 incidents of forest fires with information on days, months, coordinates, and some indexes obtained from FWI such as FFMC, DMC, DC, ISI, and burnt land area and record of meteorological conditions that occur such as temperature, wind, humidity and rainfall [5].

In previous studies many predictions have been made for forest fire prevention, including using data mining methods, such as random forest and support vector machines [5], as well as genetic algorithms [7], [8]. The method that will be used in this study is one of the machine learning methods, namely a neural network which is basically a mathematical model inspired by the biological nervous system in the brain. The advantages of neural networks when compared to other machine learning methods are in the learning process that has adaptive properties so that this method has the ability to learn from the data provided for initial training, and has a self-organization mechanism which is a neural network can make its own organization after receive information during the training period, besides that calculation on the neural network can be done in parallel and real-time [9].

Neural networks have been successfully applied in case of predictions by producing good accuracy when compared to other algorithms [10]. Neural networks have a lot of architecture in learning processes, one of which is Extreme Learning Machines (ELM), which have also been applied in many previous studies and have been shown to provide satisfactory results [11], [12], [13]. In this study ELM method will be used to predict the total area of land that will be affected by forest fires based on meteorological conditions, namely temperature, wind, humidity and rainfall, so that the ELM method is expected to provide predictive value with the best accuracy compared to the method in previous studies, so that it can contribute to the prevention of forest fire disasters. To produce the best accuracy value, it is necessary to do an architecture design on ELM to improve the performance of the method, so that this study also tests ELM parameters to find out the best parameters that will optimize ELM performance.

II. LITERATURE REVIEW

A. Extreme Learning Machine

978-1-7281-3880-0/19/$31.00 ©2019 IEEE

The Extreme Learning Machine (ELM) is one of the models of Artificial Neural Networks (ANN) that was first developed by Huang in 2004 [9]. The ELM method was developed in response to the problems found in the feed forward ANN, namely old learning speed. Huang stated two reasons that feed forward ANN has a long learning time because it uses a slow gradient-based learning algorithm to do learning and all parameters on the network are determined repeatedly using the learning method [10].

The working principle of ELM only uses one hidden layer and the process of determining its parameters is randomly generated. It is able to provide a more stable quality of accuracy with a relatively faster time compared to ordinary feed forward networks [9]. In the ELM method, the learning process is divided into 2 namely the training process and the testing process. In the ELM training process consists of a number of steps starting from initial initialization and bias, calculating the hidden layer output matrix, activating the activation function calculating the Moore-Penrose Pseudo Inverse matrix, calculating the output weight matrix value and finally calculating the predicted results. Whereas in the process of testing the sequence of steps carried out is the initialization of initial and bias weights, calculating the value of the hidden layer output matrix, activating with the activation function, and finally calculating the predicted results.

B. Multiple Linear Regression

Multiple linear regression is an extension of linear regression with the case where y is a linear function of two or more independent variables. For example y is a linear function of the variables x_1, x_2 until x_n as follows:

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + \cdots + \epsilon \quad (1)$$

The equation is very useful for matching experimental data when the variables studied are a function of two other variables. In the two-dimensional case, the "line" of regression will change to "shape". To determine the coefficients of a0, a1 and a2, the "best" value of the coefficients is determined by formula sum of square of the residuals:

$$S_r = \sum_{t=1}^{n} (y_t - a_0 - a_1x_{1,t} - a_2x_{2,t})^2 \quad (2)$$

Then the derivative of any unknown coefficients calculated are a_0, a_1 and a_2 as follows:

$$\frac{\partial S_r}{\partial a_0} = -2\sum (y_t - a_0 - a_1x_{1,t} - a_2x_{2,t}) \quad (3)$$

$$\frac{\partial S_r}{\partial a_1} = -2\sum x_{1,t}(y_t - a_0 - a_1x_{1,t} - a_2x_{2,t}) \quad (4)$$

$$\frac{\partial S_r}{\partial a_2} = -2\sum x_{2,t}(y_t - a_0 - a_1x_{1,t} - a_2x_{2,t}) \quad (5)$$

The unknown coefficients of a0, a1 and a2 can be obtained by determining the partial derivative of equation (2), (3) and (4) which equal to zero then presented in the representation of the matrix form I below:

$$\begin{bmatrix} n & \sum x_{1,t} & \sum x_{2,t} \\ \sum x_{1,t} & \sum x_{1,t}^2 & \sum x_{1,t}x_{2,t} \\ \sum x_{2,t} & \sum x_{1,t}x_{2,t} & \sum x_{2,t}^2 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} \sum y_t \\ \sum x_{1,t}y_t \\ \sum x_{2,t}y_t \end{Bmatrix} \quad (6)$$

Multiple linear regression is the most common used in forecasting research. There are two main advantages to analyzing data using a multiple regression model. The first is the ability to determine the relative influence of one or more predictor variables to the criterion value and the second advantage is the ability to identify outliers, or anomalies [11].

C. Random Forest

The Random Forest Regression is a learning method to solve regression problems by building several decision trees with outputs in the form of predictions of the average of each decision tree [12]. Random Forest was prepared to correct the shortcomings of decision tree methods that have the possibility of over fitting in the process. Random Forest runs efficiently on large databases. It can handle thousands of input variables without variable deletion. Random Forest has an effective method for estimating missing data and maintains accuracy when a large proportion of data is missing and capable of handling outlier detection problems [13].

III. METHODOLOGY

In predicting the total area of land affected by forest fires, the research methodology will be used in the form of a flowchart in Figure 1.

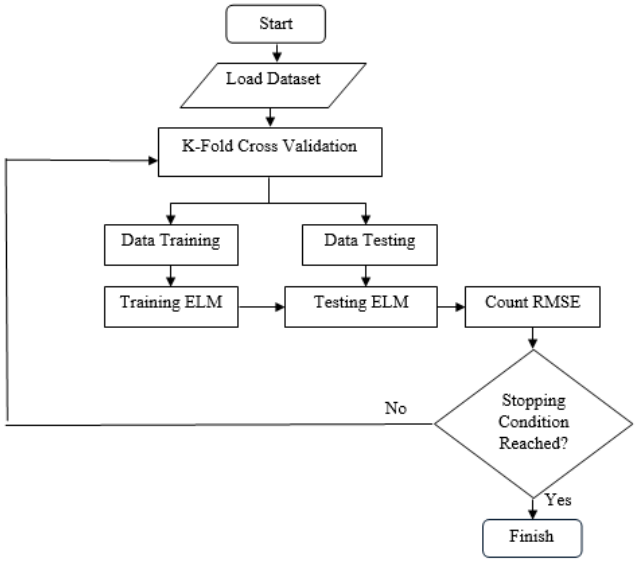

Figure 1. Research Methodology

The dataset used was obtained from Montenshino Natural Park in Portugal with a total of 517 incidents of forest fires. The dataset consists of 13 attributes and 517 data lines. The features used, namely: X, Y, month, day, FFMC, DMC, DC, ISI, temp, RH, wind, and rain. While the area feature is the value that will be predicted using ELM. Figure 2 is a sample from the dataset that will be used in this study.

	X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area
0	7	5	mar	fri	86.2	26.2	94.3	5.1	8.2	51	6.7	0.0	0.0
1	7	4	oct	tue	90.6	35.4	669.1	6.7	18.0	33	0.9	0.0	0.0
2	7	4	oct	sat	90.6	43.7	686.9	6.7	14.6	33	1.3	0.0	0.0
3	8	6	mar	fri	91.7	33.3	77.5	9.0	8.3	97	4.0	0.2	0.0
4	8	6	mar	sun	89.3	51.3	102.2	9.6	11.4	99	1.8	0.0	0.0

Figure 2. Sample of dataset

Testing is done by applying the k-fold cross-validation, with a k value that is 10. Using cross-validation to determine the number of data train and testing is done to effectively get the right amount of training data and testing. The ELM experiment in this study was carried out by comparing the Root Mean Square Error (RMSE) and computation time based on the number of neurons and hidden layers with the number 10, 20, 40, 60, and 80.

Besides that, it will also be compared based on its activation functions, namely linear, Radial Basis Function (RBF), and sigmoid activation. The purpose of this experiment is to determine the type of activation function and the number of neurons that can provide optimal results.

After making a comparison of the parameters on the ELM then the Mean Square Error (MSE) value and the best computation time will be obtained. The MSE value and computation time of ELM will be compared with the comparison method, namely: Linear Regression, Random Forest Regression, and Support Vector Regression (SVR). SVR specifically will be tested based on the kernels, namely linear, polynomial, RBF, and sigmoid kernels.

MSE is used to evaluate the regression model and to determine the level of accuracy. The smaller the MSE value is the better the regression model used. Following is the formula to calculate the MSE value:

$$MSE = \frac{\sum(y_i - \hat{y}_i)^2}{n} \quad (7)$$

y is the predicted value
$\hat{y}$ is the observed value
n is the amount of data used

IV. EXPERIMENT AND EVALUATION

The results in this chapter are comparisons of RMSE values from the proposed method, namely ELM, as well as comparison methods, namely Linear Regression, Random Forest Regression and Support Vector Regression (SVR). Comparison of RMSE values is tested based on the number of neurons in the hidden layer and each activation function.

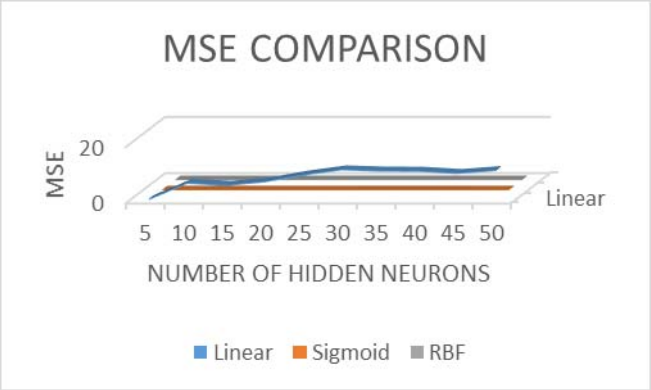

Figure 3. RMSE graph based on activation function

From Figure 3, it can be seen that MSE in the RBF activation function is more stable even though the number of neurons in the hidden layer has been changed. However, the MSE value is greater than the other activation functions. Whereas with the sigmoid activation function, the RMSE value produced decreases with the increase in neurons in the hidden layer. MSE on RBF activation is better than sigmoid activation function. The linear activation function produces the best MSE compared to other activation functions and the MSE value is fluctuating in each number of neurons tested.

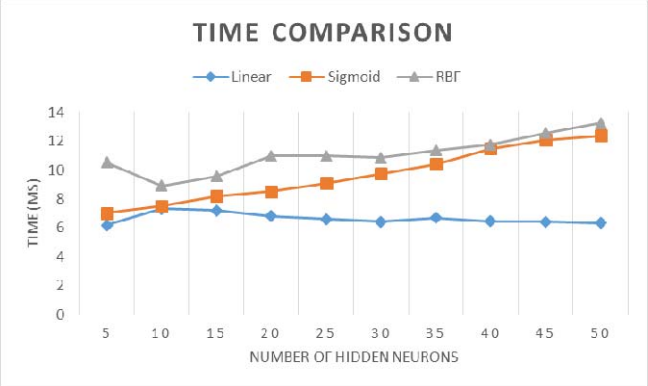

Figure 4. Graph of computation time based on activation function

The choice of activation function and the number of neurons can influence the computational time of the regression process. Based on Figure 4, the more the number of neurons, the longer the computational time needed during the regression process. The activation function with the fastest computing time is a linear activation function, while the sigmoid activation function produces the longest computation time. In Table 1, the detailed test results obtained from the ELM architecture are presented. It can be seen from Table 1 that the linear activation function with the number of neurons is 20 resulting in the smallest RMSE of 63.09511, and the fastest computation time in the number of neurons 10 with 11 ms.

Table 1. RMSE and ELM computing time based on the activation function

	Linear Activation Function		RBF Activation Function		Sigmoid Activation Function	
Number of neurons	Computation Time	RMSE	Computation Time	RMSE	Computation Time	RMSE
5	**0,334034892**	**6,183505058**	0,012982466	7,032012939	0,003537411	10,51948071
10	6,074990849	7,330405712	0,008701278	7,50528574	0,003537411	8,924210072
15	5,312619001	7,213950157	0,007377771	8,194835981	0,003537411	9,573125839
20	6,568898843	6,831562519	0,006444705	8,539766073	0,003537411	10,96976399
25	8,989307706	6,622142792	0,005986708	9,105744362	0,003537411	10,969944
30	10,86277239	6,432640553	0,005598821	9,749058882	0,003537411	10,87032159
35	10,42374124	6,696087973	0,005955867	10,40807111	0,003537411	11,36903422
40	10,35633178	6,457406282	0,006148292	11,50062382	0,003537411	11,76811755
45	9,566336831	6,426925129	0,00605397	12,09551758	0,003537411	12,5550641
50	10,62531005	6,342725754	0,005845811	12,39191055	0,003537411	13,26292038

From the results of the comparison above, it can be seen that the RMSE generated from each activation function and on a number of neurons produces values that are not much different. The difference is the computational time which increases with increasing number of neurons. But the computation time is very short, nothing reaches 1 second. The longest time is 13,26292038 milliseconds in the sigmoid activation function with 50 neurons. It can be concluded that the linear activation function is the most suitable activation function in this dataset.

Next, we will compare with the comparison methods, namely in the Linear Regression, Random Forest Regression, and Support Vector Regression (SVR) as presented in Figure 5. In the SVR method, the kernel used is the sigmoid kernel and what we will display is SVR with sigmoid kernel and RBF.

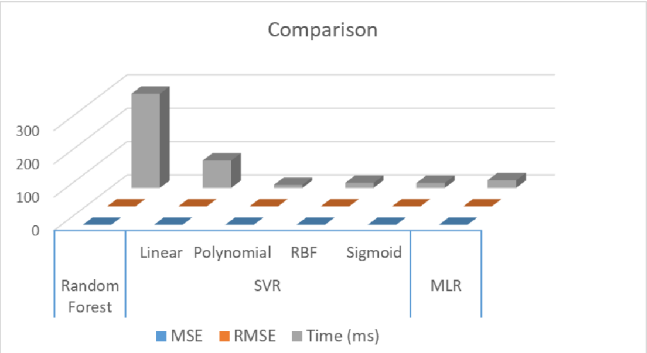

Figure 5. Comparison of MSE, RMSE and computational time with other regression methods

It can be seen from Figure 5 that Random Forest gives the worst result compared to other methods, both from the RMSE value and the computation time. The computing time of the Random Forest is 282,3 milliseconds, much higher when compared to Linear Regression, SVR or ELM. From the comparison of the regression methods above, Linear Regression is the best method because it can provide the smallest RMSE and short computing time. ELM occupies the second best after Linear Regression. The difference in RMSE between ELM and Linear Regression is not much different, Linear gives results of 62.78297 and ELM is 63.09511. Other comparison methods produce RMSE of more than 64. Linear Regression and ELM methods that use linear activation functions are proven to have good results, so it can be concluded that the data used in this study are linear.

V. CONCLUSIONS AND SUGGESTIONS

Based on the research carried out, the ELM method was able to provide predictions of forest fire datasets taken from Montenshino Natural Park in Portugal with a number of 517 forest fire incidents with information on days, months, location coordinates, and some indexes obtained from FWI such as FFMC, DMC, DC, ISI, and burned land area as well as records of meteorological conditions that occur such as temperature, wind, humidity, and rainfall. Testing of ELM parameters and architecture has been done to get the best accuracy value. The best accuracy value is obtained from error calculations using RMSE. In this study the ELM method can produce an RMSE value of 63.09511 with the computation time of 25.9 ms using an ELM parameter namely, the number of hidden neurons is 20. Based on the results of the comparison the ELM method can produce results that compete with the Linear Regression method. same. It can be concluded that the best method of solving forest fire prediction cases using this dataset is Linear Regression and ELM. In further research to improve the performance of ELM hybridization methods can be done to improve the accuracy of ELM.

REFERENCES

[1] L. Tacconi, "Forest Fires in Indonesia: Causes, Costs and Policy Implications," Center for International Forestry Research, vol. 38, no. 38, 2003.

[2] R. Niu and P. Zhai, "Study on forest fire danger over Northern China during the recent 50 years," Climatic Change, vol. 111, no. 3–4, pp. 723–736, 2012.

[3] P. Hirschberger, Forests Ablaze Causes and Effect of global forest fire. Deutschland: World Wide Fund for Nature, Berlin, 2017.

[4] M. L. Tardivo, P. Caymes-scutari, G. Bianchini, M. . Garabetti, A. Cencerrado, and A. Cortes, "A Comparative Study of Evolutionary Statistical Methods for Uncertainty Reduction in Forest Fire

Propagation Prediction," Procedia Computer Science, vol. 108, pp. 2018–2027, 2018.

[5] P. Cortez and A. Morais, "A Data Mining Approach to Predict Forest Fires using Meteorological Data," New Trends in Artificial Intelligence, pp. 512–523, 2007.

[6] J. Pinol, J. Terradas, and F. Lloret, "Climate warming, wildfire hazard, and wildfire occurrence in coastal eastern spain," Climatic Change, vol. 38, pp. 345–357, 1998.

[7] M. Denham, K. Wendt, G. Bianchini, A. Cortés, and T. Margalef, "Dynamic Data-Driven Genetic Algorithm for forest fire spread prediction ," Journal of Computational Science, vol. 3, no. 5, pp. 398–404, 2012.

[8] H. Hong, P. Tsangaratos, I. Ilia, J. Liu, A. Zhu, and C. Xu, "Applying genetic algorithms to set the optimal combination of forest fi re related variables and model forest fire susceptibility based on data mining models . The case of Dayu County, China," Science of the Total Environment, vol. 630, pp. 1044–1056, 2018..