Proceedings of the 4th ITB Graduate School Conference
Innovation and Discovery for Sustainability
July 6, 2023

Inundation Height Prediction using Machine Learning with Artificial Neural Network (ANN)

Nabila Siti Burnama [1,*], Faizal Immaddudin Wira Rohmat [2,3], Mohammad Farid [3], Hadi Kardhana [3] & Arno Adi Kuntoro [3]

[1]Graduate School of Water Resources Engineering, Faculty of Civil Engineering, Bandung Institute of Technology, Jl. Ganesa No. 10 Bandung 40132, Indonesia
[2] Water Resources Development Center, Bandung Institute of Technology, Jl. Ganesa No.10 Bandung 40132, Indonesia
[3] Water Resources Research Group, Faculty of Civil Engineering, Bandung Institute of Technology, Jl. Ganesa No.10 Bandung 40132, Indonesia
[*]Email: nabilasitiburnama@gmail.com

Abstract. Predicting the water level of inundation for the Majalaya Area is important for preventing huge losses because of floods. Because the water level increases rapidly when it rains within one to two hours, and people have time to mitigate. In this study, Machine Learning (ML) with Artificial Neural Network (ANN) method is applied to predict the inundation map in Majalaya Area. Satellite rainfall data from 2014 until 2020, distance to the nearest river, and distance to inflow from Merit DEM are the dependent variables. The inundation height from HEC-RAS extreme rainfall simulation from 2017 until 2020 is the independent variable. All data for the ANN model is divided into training and testing data for machine learning, with 75% data for training and 25% for testing. The results of training machine learning are the inundation height by ANN and inundation height by HEC-RAS have Multiple R-Square (R2) is 0.9212, and for testing data, the R2 is 0.9277. The conclusion is that the machine learning model can predict the inundation height with a high correlation coefficient result. The development of this idea and model is expected to reduce the loss of flooding and improve mitigation.

Keywords: *machine learning; ANN; prediction; mitigation; inundation height.*

1 Introduction

Floods become the natural disaster that causes losses the most. The loss can be significant property damage, business disruption, and loss of life. The damage may get worse as the climate changes over time. Climate change will continue to grow as the human population and the impending magnitude of global climate change [1,2]. Also, floods can threaten the security of society and the normal development of the economy in cities caused by extreme rainfall and the vulnerability and resiliency of the affected area. Human losses from flooding are projected to rise by 70 % - 83 % and direct flood damage by 160 % - 240 %, depending on the socio- economic scenario. Understanding river flooding and its impact is critical to reducing flood risk [1,3]. Development in flood mitigation can face the urgent challenge of climate change in the coming decades. Based on the earth's current condition, climate change has increased drastically, and the rainfall characteristics are changing, causing extreme rainfall to occur more frequently [4].

Indonesia, a country prone to natural disasters, has faced immense losses. The scale of these disasters is evident from the national disaster management board's statistics. Among the provinces, West Java stands out as the most heavily affected by natural disasters, with a record-breaking total of 1,877 incidents reported throughout the year [5]. The most natural disasters in West Java province have been recorded; 1.877 incidents were recorded throughout 2021. These numbers serve as a stark reminder of the ongoing challenges faced by Indonesia in managing and mitigating the impact of natural disasters. The frequency and severity of these events underscore the urgency of developing effective strategies and disaster response mechanisms to protect the affected communities' lives, livelihoods, and infrastructure.

A mitigation plan can reduce losses caused by a flood. In the study case, Majalaya had a short time for citizens to evacuate themselves before the flood happened. Majalaya Area is part of Bandung metropolitan cities that experiences the most massive floods occurs in West Java. Using the Majalaya case, as a study location, based on local citizens don't have enough time to evacuate themselves after rain happens. Usually, it happens one to two hours after rain. A mitigation system is needed to help people know the flooding faster to reduce the risk. So, in this study, the prediction of flood inundation height from several extreme rainfall events was carried out using four predictor variables: hourly rainfall, distance from the river, distance from inflow, and elevation using Machine Learning with Artificial Neural Network (ANN) method.

Some research has provided that using Machine Learning improves the prediction accuracy of the classic statistic model because Machine Learning can indicate the relationship in nonlinear data, as well as the correlation between predictors [6,7]. Although the logistic regression and ANN model come from the same roots in statistical regression, the standard regression must have some limits due to the amount of data required to get an accurate prediction model [6,7].

The purpose of predicting the water level is to inform the citizens of the probability of flooding, she o the government, so the people can prepare to evacuate themselves. The economic, human, and others can be prevented. So, in this modeling, from several approach variables, rainfall data is a variable factor that changes. Based on sources from the Majalaya community, which currently uses a mitigation system as a community approach [4,8].

2 Methodology

In this research, the prediction of the inundation height in the Majalaya Area model consists of some approximate such as hydrology data: hourly rainfall data, and geometric data: elevation, the distance to river inflow, and the distance to the nearest river. For more understanding, the research work describes in Figure 1 below:

2.1 Study Area

Majalaya Area is located at 107°34’E - 107°50’E and 07°02’S - 07°16’S Upstream of Citarum River Watershed, West Java province, Indonesia. Majalaya Area comprises three subdistricts such as Paseh, Majalaya, and Ibun, with the area for Majalaya Watershed being 270 km2. Citarum River is the largest and longest river in West Java. It has struggled with problems like sedimentation and overflowing rubbish [8,9].

Majalaya area experiences floods around six times yearly It takes a short time for a flood happened, and the floods that were increasing rapidly. Majalaya experience flash floods in 2018 that caused massive losses. When this flood happened, the discharge of the Citarum River was 155 m3/s, creating 5 km2 of flooding area across dozens of losses such as damaged houses, blocking the transportation route, people injured, and claiming one life [10].

2.2 Data

This study uses some hydrology satellite data and topography data.

2.2.1 Rainfall Data

For rainfall data using satellite-based rainfall, GSMaP (Global Satellite Measurement Precipitation) with a spatial and temporal resolution (0.1° x 0.1°, 1 h). Satellite GSMaP rainfall data is available in several versions, and the one used for this research is version-7 from 2014 to 2022 hourly. This satellite rainfall is downloaded from JAXA, in rainfall data with language programming R, and adjusted to Indonesia’s time zone.

2.2.2 Water Level Data

The water level of the Citarum River in Majalaya uses the results of Citarum Automatic Water Level Recording (AWLR) records managed by BBWS Citarum. The temporal data scale is every ten minutes from 2017 to 2022. However, it is assumed that the data from the results of AWLR recording are underestimated, so data correction is carried out by comparing the results of AWLR recording with manual recording. Where at the point where the AWLR is installed, there is a water gauge board, which is usually used to read the water level.

2.2.3 Topography Data

Topography is an important aspect of correct hydraulic simulation. In this study, the topography data is using a Merit Digital Elevation Model (DEM) with 5°x 5° tiles. Merit DEM is chosen because it's developed by removing some errors from past DEM data [11]. The study area was flood prone in Majalaya Area from the flood that happened in February 2018.

By utilizing the crowdsourced data, the inundation area is known as about 5 km2. The crowdsourced data will be used as the boundary for the inundation map simulation with HEC-RAS [12].

Table 1 Summary of parameters

Data	Period Time	Source
Majalaya watershed	2018	DEM (Digital Elevation Model)
Terrain data	2018	Survey Data
Flood event boundaries	2018	Survey Data
Land coverage map	2006	BIG (Geospatial Information Agency)
Satellite precipitation	2014 – 2022	JAXA
Rain gauge precipitation	2014 – 2022	BMKG (Meteorology, Climatology, and Geomorphology Agency)
Citarum river discharge	2017 - 2022	BBWSC (Citarum River Area Agency)

2.2.4 Data Preparation

2.2.4.1 Event Selection

Water level data is used to calculate river discharge using the rating curve equation published by BBWS Citarum. This formula is obtained by calculating the equation of the rating curve using historic data from 2006 – 2012.

$$Q = 3.25\ (H - 0.03)2.355 \qquad (1)$$

Where Q is discharged (m3/s), H is the water level (m). After obtaining the discharge, based on the calculation of the cross-sectional capacity of the Citarum River, the capacity is 100 m3/s [10].

The event that the discharge has selected is larger than 100 m3/s because it means that the water will be overflowing. The inundation map simulation with HEC–RAS will be based on these events.

2.2.4.2 Euclidean Distance

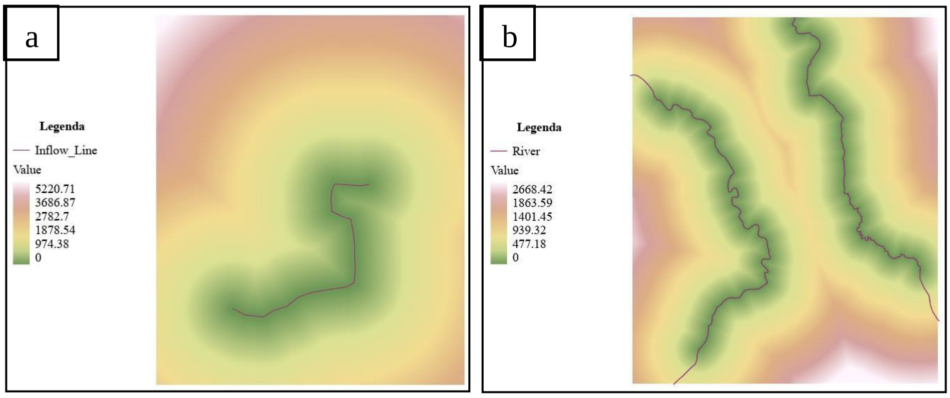

Figure 1 (a.) Euclidean Distance Results for Distance to Inflow River; (b.) Euclidean Distance Results for Distance to Nearest River.

Euclidean distance is a method to calculate the distance with a certain point. Some information from DEM that we can get is the distance from the coordinate point to the nearest river, the distance from the coordinate to the river inflow, and elevation. As shown in Figure 2, the results calculate the distance between a grid into a certain point, like the inflow river and the nearest river. The different color in Figure 2 is to inform the classification distance. The distance to the river inflow and the distance to the nearest will be the input data for the machine learning model.

2.2.5 HEC -RAS Simulation

Hydrology Engineering Center's River Analysis Software (HEC-RAS) is a hydraulic model software that the researcher widely uses for model inundation maps. This software, one of the best software for simulation inundation maps, requires some parameters like river discharge, topographic data, river geometry, Manning coefficient, and some structural aspect such as bridges and dams [12,13].

The results of the flood inundation map will be extracted to obtain the water level along with the coordinates of the inundation points. The information from the inundation map is the height of the inundation and the coordinate. The inundation height will be the output for the Machine Learning model as the inundation height from observation data.

2.3 Machine Learning

This study uses a Machine Learning model with Artificial Neural Network (ANN) method.

2.3.1 Normalizer

Normalization is carried out to reduce errors in the model, namely changing the dataset's scale to smaller but not destroying the values of the dataset. In this study, the Min-Max method is used for normalizing the dataset.

$$Normalizer = (x - \text{min})/(\text{max} - min) \quad (2)$$

Min-Max is a classic normalizing method of data that is widely used. This method has processed the dataset into a specific range (0-1) but losses negative information. The purpose of normalizing the dataset is to improve the accuracy of the Machine Learning model [14].

2.3.2 Artificial Neural Network (ANN)

A Neural Network is an algorithm that determines relationships in a data set where the processes carried out by this algorithm mimic the workings of the human brain, namely the neuron system. Therefore, the Neural Network component is described as a human neuron cell. Where a neuron consists of dendrites, a cell body, and an axon, each of these parts represents a part of the neural network, such as dendrites, which receive information from other cells, described as a model's data input layer. The cell body of a neuron plays a role in processing the information received, and this is the same as the second layer, where the model processes information. Finally, the axon, which provides processed information to other cells, is the same as the last layer, which will contain our data output.

The ANN model will be modeled using the RStudio application with the "neuralnet" library with four input units, two hidden layers, and one output unit, as shown in Figure 3. Another library, "tidyverse" is important because it has many functions to help manage the data [15].

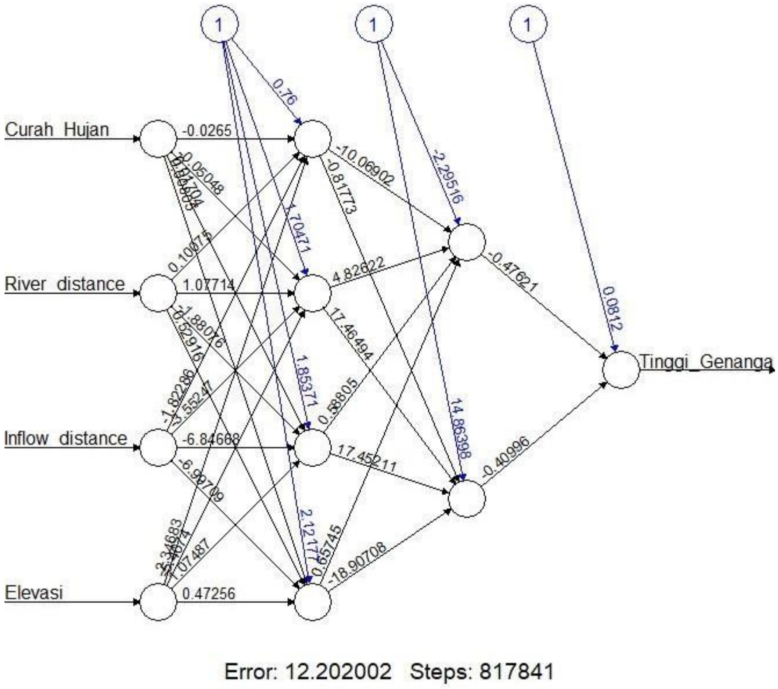

Figure 2 Machine Learning, ANN model Architecture.

For modeling machine learning, data will be divided into training and testing data. These two data ratios are 75% for training data and 25% for testing data. Training data serves as a learning medium for models to determine the effect of each input unit used.

3 Results

3.1 Training Model

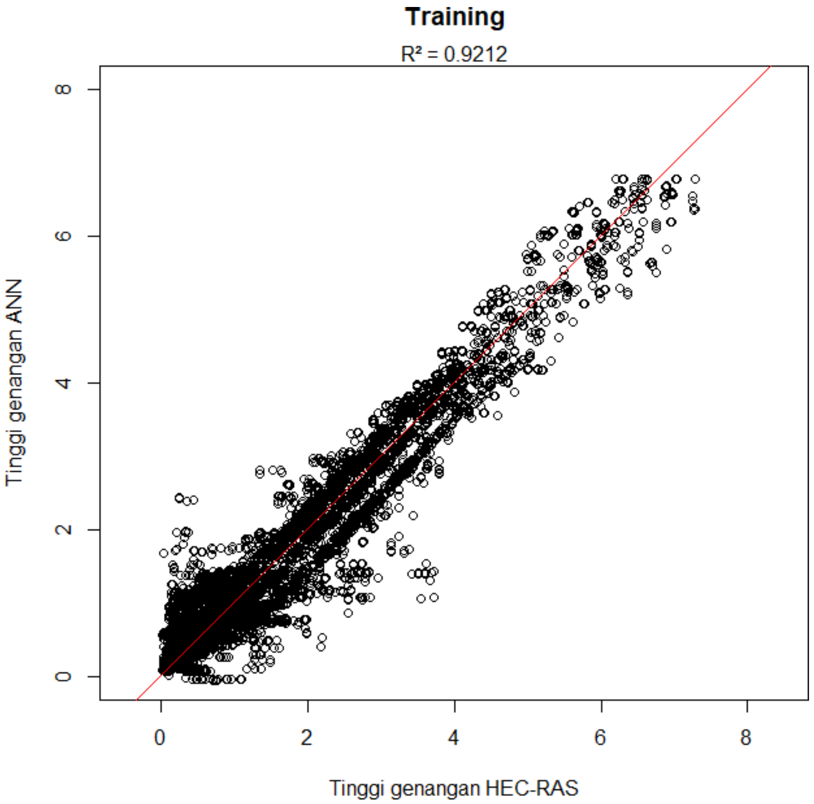

Figure 3 Inundation Height Prediction Results by ANN Model, compared with Inundation Height HEC-RAS using Training Data.

Using a neural network model, Figure 4 compares the actual inundation height with the predicted inundation height. The model prediction results are successful because the dots tend to converge on the red line. This means these two data have close values because they tend to be near the 45° line.

Furthermore, the Multiple R-Square (R2) values are calculated to measure the relationship between the two data, and the result is 0.9212. This shows that the predicted inundation height is influenced by 90% by the actual inundation height. And it provides excellent value.

3.2 Testing Model

1.1 Testing Model

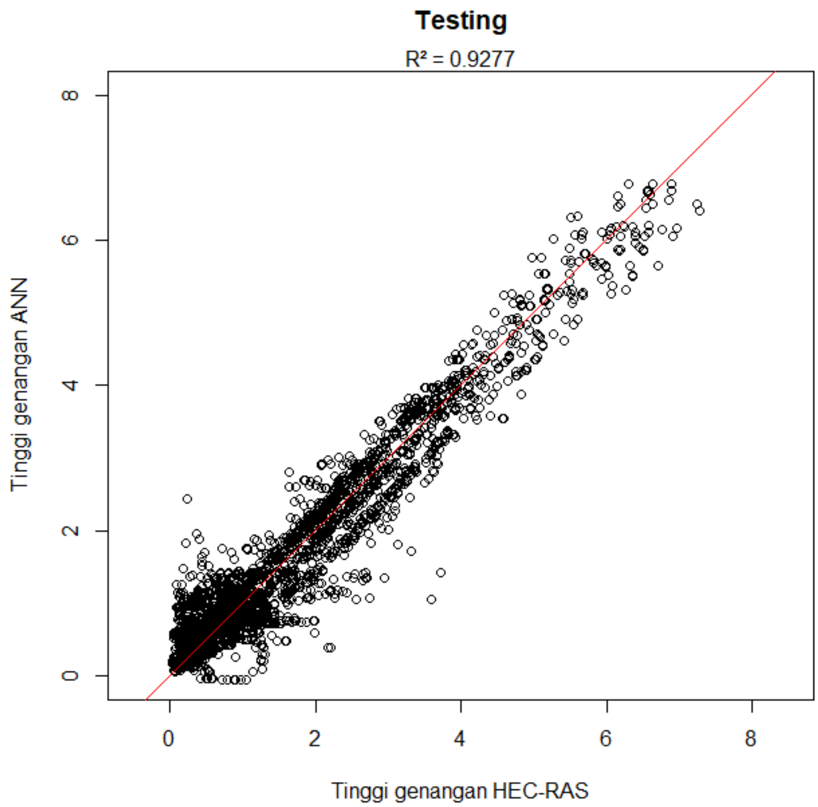

Figure 4 Inundation Height Prediction Results by ANN Model, compared with Inundation Height HEC-RAS using Testing Data.

After obtaining the prediction model, the model will try to include testing data to predict inundation height. The result can be seen in Figure 5 that the predicted inundation height results are close to the actual inundation height, so the locations of the points tend to be on the 45° line.

Then test the correlation value between the two data with the R2 value obtained 0.9277. This means the predicted inundation height is influenced by 90% by the actual inundation height. The two data have a high correlation value greater than 0.8. Then the neural network model is successful in predicting inundation height.

4 Discussion

Artificial Neural Network (ANN) model is famous among the research for more accuracy in predicting and classification. This ANN model is widely used in various fields of science. The utilization of ANN for screening and classification has proved that using ANN is more accurate than the classic statistic methods [16,17] because we know that the amount of data influences the model's capability to predict. The Machine Learning model needs a large amount of data and is inappropriate for small datasets [7].

For future research, utilizing the ANN model for making a prediction model must be developed and widely used, especially for mitigation and forecasting natural disasters. Model ANN will cut off our time for doing the calculation and simulation, as we know that disaster will came anytime [13], so predicting the future will be an efficient mitigation plan.

5 Conclusions

The results validation uses Multiple R-Square (R2) or correlation values. The results of the Machine Learning model are good if the correlation factor is higher than 80%. The neural network model enhanced by normalization provides satisfactory predictive results. For training data, the modeling results provide a high inundation value almost the same as the actual inundation level and a correlation value of 0.9212. As for the data testing model, the predicted inundation height is close to the exact inundation level with a correlation value of 0.9277. So, it can be concluded that the inundation height pre-diction modeling was successful.

In determining the number of various approaches used, it is necessary to consider the alignment between the amount of input and output data or the desired result. In this study, the amount of input data should be close because the model's results, namely the inundation map, have only sixteen events. This study is limited to the flood-prone area in Majalaya and short- range historical data (rainfall data and selected events). And it will not be good if applied to another area with a different condition because of the lack of dynamics in training data. For developing this model, some other approximate variables can be added like meteorology, geometric, and hydrology data.

Acknowledgements

This research was supported by the Bandung Institute of Technology (ITB) DTTP research grant for contract number 1354/ITI.B05/KP/2021 and ITB Research Excellence Grant (RUI) with contract number LPPM.PN-6-35- 2022.

References

[1] Munoz SE, Giosan L, Therrell MD, Remo JWF, Shen Z, Sullivan RM, et al. Climatic control of Mississippi River flood hazard amplified by river engineering. Nature 2018; 556:95–8. https://doi.org/10.1038/nature26145.

[2] Rosenzweig BR, Herreros Cantis P, Kim Y, Cohn A, Grove K, Brock J, et al. The Value of Urban Flood Modeling. Earths Future 2021;9. https://doi.org/10.1029/2020EF001739.

[3] Chen Y, Syvitski JPM, Gao S, Overeem I, Kettner AJ. Socio- economic impacts on flooding: A 4000-year history of the yellow river, China. Ambio 2012; 41:682–98. https://doi.org/10.1007/s13280-012-0290-5.

[4] Kyselý J, Beranová R. Climate-change effects on extreme precipitation in central Europe: Uncertainties of scenarios based on regional climate models. Theor Appl Climatol 2009; 95:361–74. https://doi.org/10.1007/s00704-008-0014-8.

[5] BNPB Verifikasi 5.402 Kejadian Bencana Sepanjang Tahun 2021 - BNPB 2022. https://bnpb.go.id/berita/bnpb-verifikasi-5-402- kejadian-bencana-sepanjang-tahun-2021 (accessed December 30, 2022).

[6] Nusinovici S, Tham YC, Chak Yan MY, Wei Ting DS, Li J, Sabanayagam C, et al. Logistic regression was as good as machine learning for predicting major chronic diseases. J Clin Epidemiol 2020;122:56–69. https://doi.org/10.1016/j.jclinepi.2020.03.002.

[7] Van Der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: A simulation study for predicting dichotomous endpoints. BMC Med Res Methodol 2014;14. https://doi.org/10.1186/1471-2288-14-137.

[8] Dewa Gede A Junnaedhi I, Riawan E, Suwarman R, Wahyu Hadi T, Lubis A, Joko Trilaksono N, et al. Majalaya Flood Early Warning System: A Community Based Approach. IOP Conf Ser Earth Environ Sci, vol. 71, Institute of Physics Publishing; 2017. https://doi.org/10.1088/1755-1315/71/1/012013.

[9] Alam FC, Sembiring E, Muntalif BS, Suendo V. Microplastic distribution in surface water and sediment river around slum and industrial area (case study: Ciwalengke River, Majalaya district, Indonesia). Chemosphere 2019; 224:637–45. https://doi.org/10.1016/j.chemosphere.2019.02.188.

[10] Rohmat FIW, Sa'adi Z, Stamataki I, Kuntoro AA, Farid M, Suwarman R. Flood modeling and baseline study in urban and high population,

environment: A case study of Majalaya, Indonesia. Urban Clim 2022;46: 101332. https://doi.org/10.1016/j.uclim.2022.101332.
[11] Garrote J. Free Global DEMs and Flood Modelling—A Comparison Analysis for the January 2015 Flooding Event in Mocuba City (Mozambique). Water (Switzerland) 2022;14. https://doi.org/10.3390/w14020176.
[12] Namara WG, Damisse TA, Tufa FG. Application of HEC-RAS and HEC-GeoRAS model for Flood Inundation Mapping, the case of Awash Bello Flood Plain, Upper Awash River Basin, Oromiya Regional State, Ethiopia. Model Earth Syst Environ 2022;8:1449– 60. https://doi.org/10.1007/s40808-021-01166-9.
[13] Tamiru H, Dinka MO. Application of ANN and HEC-RAS model for flood inundation mapping in lower Baro Akobo River Basin, Ethiopia. J Hydrol Reg Stud 2021;36. https://doi.org/10.1016/j.ejrh.2021.100855.
[14] Jin X, Zhang J, Kong J, Su T, Bai Y. A Reversible Automatic Selection Normalization (RASN) Deep Network for Predicting in the Smart Agriculture System. Agronomy 2022;12. https://doi.org/10.3390/agronomy12030591.
[15] Nwanganga Fred, Chapple Mike. Practical Machine Learning in R. John Wiley & Sons, Incorporated; 2020.
[16] Hand DJ. Classifier technology and the illusion of progress. Statistical Science 2006; 21:1–14. https://doi.org/10.1214/088342306000000060.
[17] Choi SB, Kim WJ, Yoo TK, Park JS, Chung JW, Lee YH, et al. Screening for prediabetes using machine learning models. Comput Math Methods Med 2014; 2014. https://doi.org/10.1155/2014/618976.