The flood prediction model using Artificial Neural Network (ANN) and weather Application Programming Interface (API) as an alternative effort to flood mitigation in the Jenelata Sub-watershed

O M Gessang1*, U Lasmingo1

1 Department of Civil Engineering, Faculty of Civil, Planning and Geo Engineering, Sepuluh Nopember Institute of Technology, Surabaya, East Java, Indonesia.

*Corresponding author's e-mail: oscar.gessang@gmail.com

Abstract. Jenelata is a sub of the Jeneberang watershed in South Sulawesi which has rainfall intensity from 2,800 mm to 4,000 mm per year on the upstream area, some of the rainfall occurs in a short period with high intensity, resulting significant rise to river water level. It potentially to cause floods downstream. Topography of the upstream area is mountainous with an average slope of ± 0.024 along the river, with flow length of 38,314 km it has a large velocity flow. Purpose of this research is to providing alternative for flood mitigation using rainfall predictions and runoff calculations, hypothetically it can reduce the impact of possible flood. This research was conducted using Artificial Neural Network (ANN) as rainfall predictor and input variable for runoff calculation using SCS method. ANN input variable will use weather prediction data from the global weather API, while SCS method will calculate maximum runoff in the catchment for the next 24 hours. Results of the rainfall prediction model get deviation of 28.81 mm and accuracy of 58% compared to the observations. Meanwhile, runoff model discharge acquires deviation of 1181.7 m³ and water level ± 0.19 m at the designated location of water level gauge.

1. Introduction
Mitigation is a series of efforts to reduce disaster risks which can be done through physical development, awareness and capacity building to face the threats of disaster (Undang-Undang No. 24 Tahun 2007, 2007). Indonesian government regulations stated that mitigation variables for disaster events must be prepared by early warning and mitigation of disaster. The sequence of steps is to observe the possible causes of disaster, analyze and observe the data, then decisions can be made based on results and analysis of the disaster which can be disseminated to take actions (Peraturan Pemerintah No. 21 Tahun 2008, 2008). This task is very difficult to do on a daily basis where natural conditions changed dynamically with very sensitive variables such as weather. Rapid development on technology in various fields of applied science has many new possibilities in terms of its application with science methods. One of these is the application of weather prediction using Artificial Neural Network (ANN) (Loucks D. P, 2017). The application of ANN with data-driven systems can significantly help to overcome several types of recurring problems. Automation is also increasingly needed these days, as it is important to be able to produce fast, reliable and efficient outputs. However, application of these technologies and systems requires availability of sufficient amount of data to be able to accessed.
Flood is one of the disasters that causes the most damage (materially and non-materially) in Indonesia [1]. According to the Indonesian Disaster Risk reports, the data show that the main cause of flood events is rainfall [2]. National Water Resources Information System encourages provision of data and information needed to mitigate flood disasters, using hydrological, hydrometeorological and hydrogeological information as information material in decision making [3]. Apart from that, the systems and technologies owned by each agency and ministry in charge are still very limited. BMKG as the Coordinator of the Information System for Hydrology, Hydrometeorology, Hydrogeology (SIH3) [4] as well as agencies / organizations with authority in climate and weather issues seems already to have, use and apply weather prediction technology and algorithms on national scale which sometimes will not be available at a specific value on times or locations. The Ministry of Public Works and Public Housing (PUPR) also has many hydrological stations, but (also) many still use daily observations, for example, the location of the research at the upstream part of the Jenelata, Gowa Regency, South Sulawesi Province. The location is quite remote and does not have (working) telemetric climatological recorder. One alternative that can be used is the use of global weather predictions that are available as a service and part of the “Big Data”. This method can be used and developed as an application for flood mitigation integrated with ANN and Machine Learning (ML).

1.1. Research objectives
The purpose of this research is to provide alternatives for flood mitigation using predictions of rainfall and runoff calculation, which hopefully can reduce the impact of possible flood events, to which the research was conducted using Artificial Neural Network (ANN) as a rainfall predictor and input for runoff calculation. ANN input variable will use weather prediction data from the global weather API. SCS method will be used to calculate maximum runoff in the catchment for the next 24 hours.

1.2. Limitations
The research is carried out through spatial analysis, hydrology and hydraulics as basic variables in the calculation of rainfall - runoff. Limitations on the model are the output will have specific values, such as variables and coefficients on the studied area.

2. Literature review
Literature used as a reference of this research are previous research and theoretical base for Climatology and Weather Prediction, ANN, ML, Deep Learning (DL), Rain - Runoff, Information systems, Decision Support System (DSS) and legislations related to objects of the research

2.1. Weather
Weather is a dynamic and complex state of the atmosphere at certain times and places. Weather can generally be expressed with rain, air temperature, cloud cover, evaporation, humidity, and wind speed from day to day [5]. The weather forecasting method has become a lot better in the past few years. Weather forecasting for 5 days is now more reliable and comparable to 2-day forecasting done 20 years ago. This is now possible by the use of technology in collecting and processing of weather data. Complex modelling of atmospheric conditions can be done using computers that are also increasingly sophisticated. Despite technological advances, accuracy in weather forecasts is still not absolute, with relatively large errors, because the weather is a complex system [6]. To be able to predict the weather, data of atmospheric conditions at the location must be obtained, such as temperature, air pressure and other atmospheric characteristics. From several weather prediction methods, an ANN model is built to predict rainfall according to the composition of atmospheric characteristics that occurs at the research site

2.2. Artificial Neural Network (ANN)
Artificial Neural Network is a set of an algorithm works on a (virtual) machine designed to work like the neural of the human brain, that it can solve certain kind of problems with specific functions. Information and data stored are obtained through the learning process [7]. The ANN is built of neurons, connecting each other by paths called synaptic, thus, forming a network of algorithms with different
paths depending on the value and weight of each synaptic. Each neuron will process input data to specific information as targets.

The use of ANN in predicting weather has been already done by many researchers such as [8], [9], [10], [11], [12]. They are looking for relationships between weather, rain and methods to get (more) accurate predictions. With a combination of sensitive variables, a weather - rainfall prediction is more non-linear in nature and (very) complex to be numerically calculated. Rainfall modelling is one of the most difficult to predict accurately, so that detailed knowledge of atmospheric condition is very important. However, it will still be difficult to predict and observe accurately. The model must be made using simple parameterization to effectively account for processes that can arise as a result of the limitations of the forecast model and other errors in observation [11]. The approach using the ANN method implemented simple, quick, and flexible problem solving. It is also easily adaptable to new environments [12]. ANN has many advantages in terms of time, cost, and ease in production. It can also be integrated with automation systems such as automatic data acquisition systems so that it can be implemented into a real-time forecasting system [8]. The advantages of ANN lie on its ability to process data flexibly in a relatively short time of reproduction and to calculate non-linear relationships between input and output variables. This makes ANN worth to be considered as a tool to estimate long-term runoff in real-time [9].

2.3. Rainfall - runoff
Precipitation or rainfall occurs from the product of evaporation and transpiration processes. It will undergo a process of infiltration, percolation, where some part of the rainfall will become overland flow and the other part will flow (underground) and get infiltrated. The amount or composition of surface runoff and infiltrated flow is affected by topography characteristic of the area. Surface runoff can be divided into 2 (two) groups: sheet runoff and surface runoff. The flow of water that enters the water flow system becomes a stream flow [13].

The value of surface runoff is affected by rainfall, vegetation, land cover and presence of water storage infrastructure / buildings. Calculation method using Curve Number (CN) is based Soil Conservation Service (SCS) where rain runoff (Pe) is a function of cumulative rain (P), land use, soil type and humidity through Maximum Retention (S), and Initial Abstraction (Ia) [14]. Soil ability to store maximum water depends on the characteristics of the soil in the watershed symbolized by the Curve Number (CN). The value is varying from 100 on waterlogged surfaces to 30 on permeable surfaces with high infiltration values.

\[
P_e = \frac{(P - Ia)^2}{P - Ia + S}
\]

\[
CN = \frac{\sum A_i \cdot CN_i}{\sum A_i}
\]

\[
S = \frac{25400 - 254 \cdot CN}{CN}
\]

\[
Ia = 0.2 \cdot S
\]

The SCS method calculates moisture condition of a soil (and its initial water content condition). This value will be calculated based on the amount of rain 5 days in advance. Indonesia has moderate climate usually compatible with CN on Antecedent Moisture Condition (AMC) II type. Other types of AMC can be calculated from CN AMC II using equation (5) or (6) [15].
\[ CN_{amic1} = \frac{4.2 \cdot CN_{amic2}}{(10 - 0.58 \cdot CN_{amic2})} \] 
\[ CN_{amic3} = \frac{23 \cdot CN_{amic2}}{(10 - 0.13 \cdot CN_{amic2})} \] 

Rating Curve equation on the Water Level Gauging Station.

\[ Q = 829.3(h - 0.614)^2 \] 

3. Research methods

3.1. Location, data acquisition and processing

Jenelata as a part of the Jeneberang watershed in South Sulawesi is located at -5.27S, 119.57E to -5.34S, 118.83E. This sub-watershed has significant value of rainfall (2800 - 4000 mm/year) at the upper reaches [16]. Jenelata river is dendritic on flow pattern with mountainous area at the upstream area. The watershed produces an average slope of 0.0238. Length of the main river is 38.31 km, with difference on elevation of ± 912.64 m from the highest point to the outlet. It has large potential for erosion upstream and sedimentation downstream. Downstream Jenelata meets the Jeneberang river which flows into the Makassar Strait. Jenelata is influenced by wet tropical climate, the west monsoon and east monsoon with normal temperatures of 24°C to 33°C. The Jeneberang watershed development guide document states that the watershed is included as an area that is prone to disasters, including soil erosion, flooding and forest fires [17].

The rainy season occurs in November to April and the dry season starts in May until October [17]. The maximum rainfall recorded at the upstream area is 900 mm (28 days) and 600 mm (20 day) in the downstream area. The Lower Jeneberang River Flood Control Project report [16] states that the average rainfall intensity in one hour is 35% of the total daily rainfall at each rain post, and 90% of daily rainfall has a mean rainfall period of 4 to 6 hours.

![Figure 2 River Basin Elevation](image)

Climatological data used in the ANN model are divided into 2 groups, namely historical climatology data in the modelling phase (training - validation) and API climatological prediction data in the forecasting phase.
Rainfall data are used in the modelling phase, the amount of which is accordance with the amount of climatological data time and location (closest). The total number of data is 1848 data: 1663 for training data and 185 for testing the model. The number of data is fairly small and, from the analysis and observations, problem comes from the data recording station (or devices). A lot of data of the rainfall station at the research location are incomplete, that it makes rain distribution area processing cannot be done, and the only complete data recording (found and) available at the research area is at Limbunga.

![Data type and processing](image)

**Figure 3** Data type and processing

Other variables used in order to calculate runoff are obtained from spatial data processing, rain prediction results generated by ANN data source models in the form of Digital Elevation Model (DEM), and land cover data obtained from the Pompengan - Jeneberang river basin management authorities (BBWSPJ).

![Land Use Map](image)

**Figure 4** Land Use Map

| Land Use          | Area          |
|-------------------|---------------|
| Natural Forest    | 83,663,673.30 |
| Mix Fields        | 69,931,400.62 |
| Swamps            | 3,862,647.12  |
| Shrubs            | 67,889,703.90 |
| River             | 1,802,729.45  |
| Grass (open field)| 2,036,172.56  |
| Fields            | 2,000,926.64  |

**Table 1** Land use table

3.2. **Modelling**

ANN model is built on Python. Data inputs are using climatology historical data within the time frame as rainfall data. The total data are then split into 2 groups: 80% of training data and 20% of validation
and test data. Trained ANN model is a result of variable for weight on each input node (neuron) stored in a file (with .h5 extension). This file can be used to make predictions.

The training process is carried out multiple times to get good accuracy predictions with the smallest deviation using training and validation data. Performance of the model is calculated using test data and measured with the value of Root Mean Square Error (RMSE).

After the model training result is considered having good performance and accuracy, rainfall prediction is carried out using input from extracted global weather variable data, that these data are acquired from API request on darksky.net (Apple Inc’s). The acquired data are processed as input on ANN model (.h5). The output of this model’s process is the prediction value for tomorrow rainfall.

![Figure 5 ANN Modelling Flow](image)

Rainfall predictions from the ANN are stored on the database, then they are used as input on the runoff calculation using SCS method. Distribution of hourly rainfall intensity is calculated using Modified Mononobe’s method with 4 to 6 hours of rainfall distribution time [16] to calculate estimated hourly discharge. Runoff calculation variables such as CN are extracted from spatial data using soil type and land use. Other variables like AMC’s are calculated using previous 5-day rainfall. This step of calculation produces the Maximum Retention (S), Initial Abstraction (Ia) of the location, rainfall depth (Q mm) and volume (Q m3).

3.3. Testing and validation
Validation process is carried out on each of the models. In the ANN model, validation process is conducted on testing phase because on the training phase it already has a validating process to measure the performance on the model’s training. The predicted rainfall is validated (or tested) by comparing prediction values to recorded data at given period of time. Meanwhile, at the runoff calculation using SCS, validation process is carried out using rainfall of a specific date / time given as input on the calculation, and compares the result with the value of recorded water level in Jenelata outlet.

4. Result and discussion

4.1. ANN modelling
The ANN model is built on Python, an architecture of the network (figure 6) consisting of 1 input layer for training data (x1) and target data (x2), 4 hidden layers, and 1 output layer (rain prediction).
Table 2 ANN Architecture

| Layer | Node | Activation | Type |
|-------|------|------------|------|
| 1     | 6    | RELU       | Input|
| 2     | 48   | RELU       | Hidden|
| 3     | 72   | RELU       | Hidden|
| 4     | 36   | RELU       | Hidden|
| 5     | 6    | RELU       | Hidden|
| 6     | 1    | RELU       | Output|

Data preprocessing phase is carried out to re-format the data to meet the Python ANN modelling requirement. Then, data are divided into 2 groups: the training data (80%) and the validation-test data (20%).

Table 3 Data training format

| No | Month | Temp (°C) | Humid (%) | Pressure (mmHg) | Wind (km/h) | Clouds (%) | Recorded Rainfall Data |
|----|-------|-----------|-----------|-----------------|-------------|------------|------------------------|
| 1  | 10    | 35.2      | 81.71     | 1076.1          | 5           | 40         | 0                      |
| 2  | 10    | 35.2      | 81.71     | 1076.1          | 4           | 40         | 0                      |
| ...| ...   | ...       | ...       | ...             | ...         | ...        | ...                    |
| 1663| 10    | 26        | 60.58     | 1076.4          | 5           | 20         | 0                      |

Result of training stage on an ANN modelling.

Train on 1663 samples, validate on 185 samples
Epoch 1/5000
1663/1663 [==========] - 0s 227us/step - loss: 247.4028 - accuracy: 0.3067 - val_loss: 25.3907 - val_accuracy: 0.5946
Epoch 2/5000
1663/1663 [==========] - 0s 100us/step - loss: 144.4335 - accuracy: 0.3957 - val_loss: 25.2943 - val_accuracy: 0.5946
... Epoch 5000/5000
1478/1478 [==========] - 0s 114us/step - loss: 2.25 - accuracy: 0.7835 - val_loss: 5.3380 - val_accuracy: 0.5216

Figure 7 ANN training Results

Validation data for the ANN are 20% of the total training data, while in network testing used 185 data, which then cross-validated with the recorded field data.

Table 4 ANN Testing

| No | Month | Temp (°C) | Humid (%) | Press (mmHg) | Wind (km/h) | Clouds (%) | Recorded Rainfall | Predicted Rainfall |
|----|-------|-----------|-----------|--------------|-------------|------------|-------------------|-------------------|
| 1  | 12.00 | 29.71     | 96.25     | 1017.00      | 4.00        | 76.00      | 30.00             | 21.206234         |
| 2  | 12.00 | 32.00     | 87.67     | 1017.00      | 5.00        | 56.00      | 29.00             | 17.563517         |
| ...| ...   | ...       | ...       | ...          | ...         | ...        | ...               | ...               |
| 185| 7.00  | 31.00     | 74.75     | 1012.00      | 3.00        | 40.00      | 1.00              | 6.478799          |

Parameter of the training model is a learning rate = 0.0002, loss = ‘mean squared error’, optimizer = ‘nadam’, metrics = ‘accuracy’, dan epoch = 5000, resulting Loss: 2.25, Accuracy: 64.72.
Model testing is similar to validation stage, and the difference is that testing stage is done by entering the value of a new weather variable as input which has never been seen by the model on the training phase or validating phase. Prediction results are then compared with the field recorded data. This test is conducted using 193-day rainfall data, the result of the predictions are (seem to be) less desirable with research expectation, but it is predicted due to quality and lack of data availability in the training model process that will lead the model to fail in predicting the value of rainfall with large intensity or extreme values.

In this case, for some climatological data that should produce extreme rainfall (large values), the model only produces predictions with large deviation values. Meanwhile, for the prediction of average rainfall (<100 mm), the model can make a pretty good prediction with RMSE of 1.45, deviation of 28.81, and the correlation coefficient between rainfall of 0.6. Analysis for this problem is the lack of data used on training phase, especially on high intensity rainfall.

4.2. Runoff modelling

Runoff calculation is using spatial variables in the proses. To be able perform runoff calculation, catchment area variables are required by calculating spatial data using processing software Quantum GIS (QGIS).

| Variables       | Value  |
|-----------------|--------|
| Area (Km²)      | 231.19 |
| Area (m²)       | 231187254.42 |
| River (Km)      | 38.31  |
| Upstream (masl) | 955.44 |
| Downstream (masl)| 42.80 |
| Slope (%)       | 2.38   |

Calculated CN of the area is 62.70, runoff calculation uses SCS method with AMC from 5-day rainfall value, while for the rain duration assumed to be 4 hours long distributed using the Mononobe method.
Testing phase of the model is conducted using rainfall data at certain period of time, calculated with model, and compared with water level measurement in Jenelata water level gauge downstream.

Calculated value (table 6) is calculated using rainfall data from January 22, 2019 as an input, where precipitation value recorded is 364 mm and maximum water level in the water level station gauge record is 4.4 m. Results of maximum volume on daily rainfall calculated using SCS rational method eq. (1), (2), (3), (4) are 53,119,118.21 m$^3$ and 14,755.31 m$^3$/s.

| Location | Precipitations | CN | S (mm) | Ia (mm) | Q(mm) | V(m3) | Q(m3/s) |
|----------|----------------|----|--------|---------|--------|-------|---------|
| Jenelata | 364.00         | 62.70 | 151.10 | 30.22 | 229.76 | 53,119,118.21 | 14,755.31 |

Distributed rainfall using modified Mononobe’s method (table 6) gets total runoff volume of 46,270,478.52 m$^3$, and maximum discharge is at the 3rd hour for 38,539,168.52 m$^3$, $\Delta Q = 6,844,668.44$ m$^3 \sim 87\%$.

| Hour | Prediction (mm) | Total (mm) | CN | S (mm) | Ia (mm) | Q(mm) | V(m3) | Q(m3/s) |
|------|-----------------|------------|----|--------|---------|--------|-------|---------|
| 1    | 33.28           | 33.28      | 79.74 | 64.52 | 12.90 | 4.89 | 1,130,982.00 | 314.16 |
| 2    | 59.60           | 92.89      | 79.74 | 64.52 | 12.90 | 19.61 | 4,532,861.60 | 1,259.13 |
| 3    | 229.31          | 322.19     | 79.74 | 64.52 | 12.90 | 166.70 | 38,539,168.52 | 10,705.32 |
| 4    | 41.81           | 364.00     | 79.74 | 64.52 | 12.90 | 8.94 | 2,067,466.40 | 574.30 |

Hypothetically, the output of the calculation can be compared to the value of Principal Hydrological Gauging Stations for Flood Warning and Evacuation of Jeneberang river [18], with such a large number of discharges (table 7), the authorities should have issued an early warning within 24 hours before it happened.

![Principal Hydrological Gauging Stations for Flood Warning and Evacuation](image)

**Figure 11** The estimated discharge will provide a flood prediction based on Principal Hydrological Gauging Stations for Flood Warning and Evacuation of Jeneberang river.

To validate the hypothesis and calculation, the runoff model is validated with water level on the observation point data during the rainy period of January 22, 2019 with 364 mm of rainfall intensity.

| Location | Observation | Model |
|----------|-------------|-------|
|          | Discharge (m$^3$) | WL(m) | Discharge (m$^3$) | WL(m) | $\Delta$WL (m) |
| Jenelata | 12,076.14 | 4.43 | 10,705.32 | 4.20 | 0.1208 |

**Table 8** Evaluation for the calculation and field observation shows that water level within the time frame has deviation of 0.12 m and total discharge of 1.370,82 m$^3$.
5. Conclusion
From the research, there are several conclusions which can be drawn:
1. Result on the rain prediction test found 28.81% of prediction error with RMSE of 1.45, and a correlation coefficient between rainfall is 0.6. This result is considerably reasonable as the model is trained with (relatively) small amount of data.
2. Significant value of deviation on extreme rainfall prediction is likely to happen due to the lack of available data used in the learning process.
3. Differences in runoff calculations using the daily SCS method and model occur with a standard deviation of 0.13. This is likely to occur because the model calculates daily rainfall distribution in 4 to 6 hours and AMC.
4. Using (decent amount of) good quality of data will make models produce better precision on the rainfall prediction.
5. Application of the ANN model can be used as an effective and efficient alternative in the decision-making process along with the numerical model.
6. Hopefully, future research can be carried out with (more) decent amount of quality data and (maybe) using other methods to make (more) applicable models that can be used in disaster mitigation as a system.

References

[1] Aldrian E K M 2011 Adaptasi dan Mitigasi Perubahan Iklim di Indonesia (Jakarta: Pusat Perubahan Iklim dan Kualitas Udara, Deputi Bidang Klimatologi, Badan Meteorologi dan Geofisika)
[2] Asdam C 2014 Hidrologi dan Pengelolaan Daerah Aliran Sungai, cetakan ke-6 (Yogyakarta: Gadjah Mada University Press)
[3] Badan Nasional Penanggulangan Bencana 2016 Risiko Bencana Indonesia (RBI) (Jakarta: Badan Nasional Penanggulangan Bencana)
[4] BBWS Pompengan Jeneberang 2015 Pola Pengelolaan Wilayah Sungai Jeneberang (Makassar: BBWS Pompengan - Jeneberang)
[5] Beven K. 2011 Rainfall-runoff modelling: the primer. In B. K. J Rainfall-runoff modelling: the primer (Oxford: John Wiley & Sons)
[6] Campolo M A S 2003 Artificial neural network approach to flood forecasting in the River Arno Hydrolog Sci J 48 p 381-98
[7] Chow V T. 1988 Applied Hydrology (New York: Mc.Graw Hill Book Company)
[8] Haykin S. 2009 Neural Networks and Learning Machines 3rd Edition (Ontario: Pearson Education Inc.)
[9] Iseh A J, W. 2013 Weather forecasting models, methods and applications Int. J. Eng. Res. 2 12 p 1945-56
[10] JICA 1980 Lower Jeneberang River Flood Control Project (Jakarta: Japan International Cooperation Agency)
[11] JICA 2005 The Study on Capacity Development for Jeneberang River Basin Management (Jakarta: Japan International Cooperation Agency)
[12] Krasnopol’sky V M 2013 The Application of Neural Networks in the Earth System Sciences, Neural Networks Emulations for Complex Multidimensional Mappings (Dordrecht: Springer)
[13] Loucks D P V B 2017 Water resource systems planning and management: An introduction to methods, models, and applications (Delft: Springer International Publishing AG)
[14] Peraturan Pemerintah No 21 Tahun 2008 2008 Penyelenggaraan Penanggulangan Bencana (Jakarta: Pemerintah Republik Indonesia)
[15] Peraturan Presiden No.88 Tahun 2012 2012 Kebijakan Pengelolaan Sistem Informasi Hidrologi, Hidrometeorologi, dan Hidrogeologi pada Tingkat Nasional (Jakarta: Pemerintah Republik Indonesia)
[16] Sharma A. K. B. (2014) Adaptive Predicting of Weather Forecasting with ANN using Back Propagation Technique *Int. J. Latest Trends Eng.* 4 297-300

[17] Toth E. 2009 Classification of hydro-meteorological conditions and multiple artificial neural networks for streamflow forecasting *Hydrol. Earth Syst. Sci.* 13 p 1555–66

[18] Undang Undang No 17 Tahun 2019 2019 *Sumber Daya Air* (Jakarta: Pemerintah Republik Indonesia)

[19] Undang-Undang No. 24 Tahun 2007 2007 *Penanggulangan Bencana* (Jakarta: Republik Indonesia)

[20] US Army Corps of Engineers 2000 *Hydrologic Modeling System HEC-HMS, Technical Reference Manual* (Davis CA: Institute for Water Resources, Hydrologic Engineering Center)