IoT based real-time monitoring system of rainfall and water level for flood prediction using LSTM Network

A A M Faudzi[1,2], M M Raslan[2], N E Alias[3]
[1] Centre for Artificial Intelligence and Robotics (CAIRO), Universiti Teknologi Malaysia, Jalan Sultan Yahya Petra, 54100 Kuala Lumpur, Malaysia
[2] School of Electrical Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia
[3] Centre for Environmental Sustainability and Water Security (IPASA), School of Civil Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia

Corresponding author: athif@utm.my

Abstract. Floods in recent years have frequently resulted in environmental, economic, as well as loss of human life. People are less aware of incoming floods if there is no early warning system. This proposal outlines the design of a monitoring system to obtain real-time data on rain gauge and water level. The monitoring system is based on IoT via a GSM network to provide real-time data cloud and dashboard display on Grafana platform. The rainfall forecasting model used Long Short-Term Memory (LSTM) networks to predict future rainfall and water level values which could cause floods. The result was experimented with using historical data since the current data of the monitoring system is insufficient yet to make an accurate prediction. The main findings of the research are the predicted values of streamflow and rainfall for historical data, also water level and rain gauge for new data. The primary result was experimented with using historical data on two rainfall stations and one streamflow. Also, the primary result was experimented with using new data on two water level stations and one rainfall. The forecasting method that applied LSTM showed high accuracy of the result reaching more than 90%. Based on these results, the system can be used as a non-structural solution to alleviate the damage caused by urban floods.

1. Introduction

Floods are one of the most prevalent and destructive natural dangers in the world [1] . Floods in previous years have been the most expensive calamities in terms of property damage and human lose [2]. More than 15.5 billion euros in damage was inflicted by the Arno River floods in Italy [3]. Between 1989 and 1999, floods in the United States claimed at least 988 lives and caused economic damages of around 4.5 billion dollars [2]. In Malaysia, Floods are one of the most common natural disasters, occurring virtually every year, particularly during the monsoon season [4]. The last flood happened in 2021, floods caused by rivers flowing into the mainland have inundated many areas, ruined buildings, blocked off important highways, and affected the provision of basic services such as water, food, and health care. According to [5] more than 18,000 families have affected by this flood. The effects of floods could be mitigated by having a flood prediction using flood monitoring system data that allows the residents to be informed quickly and efficiently. There are many issues that flood control systems encounter regarding accurate and timely forecasting of floods. In order to give a reliable prediction, it is essential to have a flood

prediction with flood monitoring system that are both accurate, timely, and provide an early warning. The objects of this study are to design and develop IoT flood monitoring system based on water level sensors and rain gauge sensors to collect accurate data. Additionally, to establish a forecasting model based on Long Short-Term Memory (LSTM) networks for early flood prediction.

This study proposes a data-driven flood prediction model using LSTM based on the accurately collected data that we have from the flood monitoring system. Section 2 introduces the structure of this work Section 3 discusses the methodology and the model design for the proposed experiment. Experimental results are shown in section 4 and lastly concluding remarks are contained in section 5.

2. The structure of the study

While conventional early warning systems may be appropriate in some circumstances, modern technology such as the Internet of Things is critical for real-time data acquisition from a variety of sources, data processing, and warning information distribution to people who are likely to be impacted by a flood before it occurs. Internet of Things (IoT) refers to the connection of physical devices, cars, buildings, and other items embedded with electronics, sensors, actuators, communication protocols, and software that collect, share, store, analyze, and process data. The structure of IoT is based on five main components which are the things or device (sensor nodes), field getaway, cloud, storage, analytic [6].

Machine learning (ML) is a subfield of artificial intelligence (AI) that is used to infer regularities and patterns. It enables easier implementation with low computation costs, as well as fast training, testing, validation and evaluation while maintaining high performance in comparison to physical models [7]. The limitations of physical-based and statistical models discussed in[8] promote the use of advanced data-driven models such as machine learning. Promising data-driven prediction models employing ML are quicker to develop with minimal inputs [9]. ML techniques have demonstrated their ability to outperform conventional flood forecasting approaches with an acceptable level of accuracy during the last two decades [9]. In these literatures [9, 10] they described in detail the ML modelling methodology and flood modelling technique and illustrate the basic flow for building the ML model.

Long Short-Term Memory was introduced by Hochreiter and Schmidhuber in 1997 [11]. LSTM is a deep learning technique which is a subfield from ML [12]. LSTMs are a special kind of Recurrent Neural Network (RNN) that has been developed to overcome the drawback of RNN in terms of the vanishing gradient problem. LSTM networks are the most commonly used variation of Recurrent Neural Networks [12]. Graves & Schmidhuber in [13] first described the LSTM architecture that is most frequently used in the literature shown in figure [1]. The memory cell and the gates are two important critical components of LSTM [14]. The information coming to the memory cell could be manipulated by the input gates and forget gates between one time-step and the next. Information may be preserved over many time steps because of the gating structure, and gradients can flow over many time steps as well. Thus, LSTM could avoid the vanishing gradient problem that happens in the majority of RNN models.

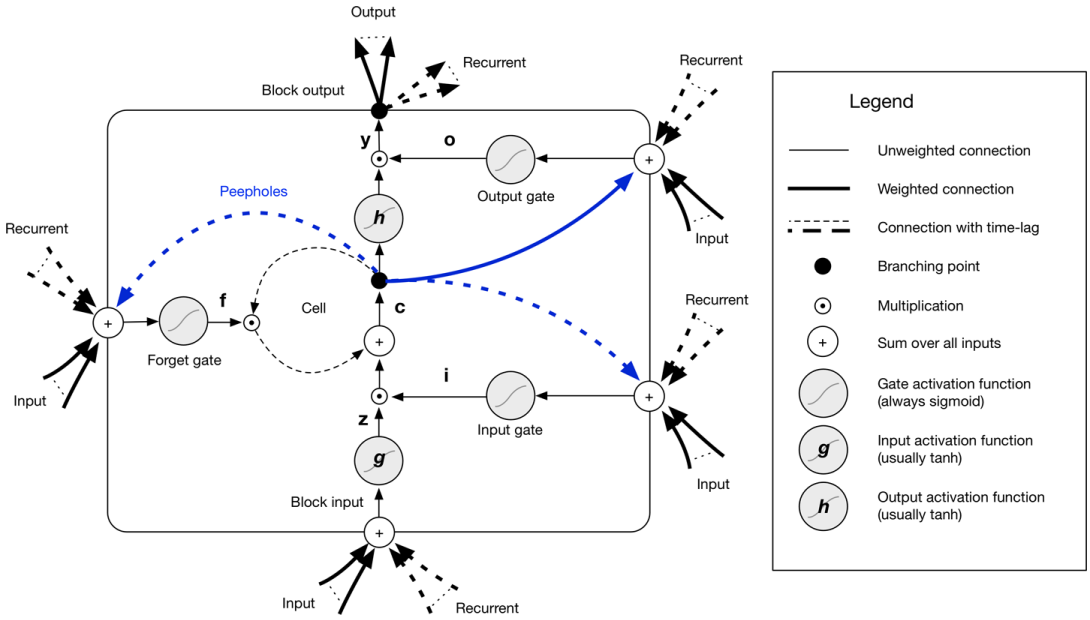

Figure 1. LSTM architecture.

The most important component of LSTM unit is the memory cell. It is an essential concept since it enables the network to retain its state over time. LSTM unit has three gates, input gate, forget gate and output gate. Protecting linear unit against misleading signals is the primary function of three gates. A memory cell's state can be changed or blocked using the input gate. Using the forget gate, a memory cell may either recall or forget how much data it has stored in its current state. At the LSTM's output, the output gate either shows or hides the contents of the memory cell. The output of the LSTM block is recurrently connected back to the block input and all of the gates for the LSTM block. An LSTM has sigmoid activation functions for [0, 1] limitation in its input, forget, and output gates. Typically, the LSTM block's input and output activation functions are tanh. Here the equations present the vector formulas for a LSTM layer forward pass [12]:

$$\begin{aligned}
\boldsymbol{z^t} &= \boldsymbol{g(W_z X^t + R_z Y^{t-1} + b_z)} && \textit{block input}\\
\boldsymbol{i^t} &= \boldsymbol{\sigma(W_i X^t + R_i Y^{t-1} + P_i \odot C^{t-1} + b_i)} && \textit{input gate}\\
\boldsymbol{F^t} &= \boldsymbol{\sigma(W_f X^t + R_f Y^{t-1} + P_f \odot C^{t-1} + b_f)} && \textit{forget gate}\\
\boldsymbol{C^t} &= \boldsymbol{i^t \odot Z^t + F^t \odot C^{t-1}} && \textit{cell state}\\
\boldsymbol{o^t} &= \boldsymbol{\sigma(W_o X^t + R_o Y^{t-1} + P_o \odot C^t + b_o)} && \textit{output gate}\\
\boldsymbol{y^t} &= \boldsymbol{o^t \odot h(c^t)} && \textit{block output}
\end{aligned}$$

where $\boldsymbol{X^t}$is the input vector at time $\boldsymbol{t}$, the $\boldsymbol{W}$ are rectangular input weight matrices, the $\boldsymbol{R}$ are square recurrent weight matrices, $\boldsymbol{p}$ are peephole weight vectors and $\boldsymbol{b}$ are bias vectors. Functions $\boldsymbol{\sigma}$, $\boldsymbol{g}$ and $\boldsymbol{h}$ are point-wise non-linear activation functions: logistic sigmoid $\frac{\mathbf{1}}{\mathbf{1}+\boldsymbol{e}^{-\boldsymbol{x}}}$ is used for as activation function of the gates and hyperbolic tangent is usually used as the block input and output activation function. The point-wise multiplication of two vectors is denoted with $\boldsymbol{\odot}$.

3. Methodology

3.1 Study area

In this project, two case study has been chosen for our modelling. The reason of choosing other two case study is that the data of the monitoring system in not sufficient to do prediction. The first case study from the hydrological stations in JB and the second case study is the location of our developed monitoring system.

The first dataset (historical data) was provided by the Official Web of Public Infobanjir. Infobanjir, a centralized database system, was developed in the year 1999 and started to be operated in early 2000. The Infobanjir system works by collecting real-time rainfall and water level data from nearly 500 hydrological stations across the country. Hydrological data from each station is transmitted to the Telemetry Database/Servers in each state and then transmitted to Infobanjir. In 2011, the Infobanjir system was modified further by developing a new web-based and renamed publicinfobanjir system. The newly developed publicinfobanjir websites were focusing on providing flood warning information to the public by reducing technical information and incorporating the latest media such as Facebook, Twitter, and RSS. Important information such as evacuation records, flood status, rainfall and river water level alerts are displayed in an interactive manner. Moreover, the database system has been enhanced and developed by applying the latest technology, which is capable of receiving and processing data in real time over a short period of time. The dataset is collected from three different stations between the 1st of June 2010 and until 1st of Dec 2012. The historical data is two Rainfall data (Site 1836001 RANCANGAN ULU SEBOL and Site 1737001 SEK. MEN. BKT. BESAR at KOTA TINGGI at JOHOR) and one streamflow data (Site 1737451 SG. JOHOR at RANTAU PANJANG). The following Figure [2] shows the location of these stations.

All three stations are installed in the Johor River Basin. The average annual temperature is about 26 °C [15]. The average annual rainfall of the river is 2500 mm/year. The mean annual streamflow at Rantau Panjang station is 37.7 m3/s. The climate in the Johor River is a tropical monsoon climate, divided into the northeast monsoon (November–February), and the southwest monsoon (May–August). Flooding events frequently occur in December, when the highest rainfall and peak streamflow are recorded. The Johor River covers four districts of Johor State: Kota Tinggi, Kluang, Kulai Jaya, and Johor Bahru. It is estimated that there were roughly 300,000 people and 70,000 families living in the Johor River Basin in Malaysia in 2010 [16]. This study collected hourly streamflow data from the station and hourly rainfall data from the two gauges.

The location of our study (new data) is in University Teknologi Malaysia (UTM). UTM campus is located in Johor Bahru in the south of Malaysia. The rain gauge is installed on an open area and at ground level, preferably site with no steep drop-off on the windward side. The sensor should not be sheltered by obstructions such as trees and buildings that may prevent rainfall from being collected by the rain gauge. Water level sensor is installed at potential flooding area particularly location which has high possibility to experience flash flood (fast rise and overflowing of river water level). There are several potential areas for the water level sensor node installation. They are at upstream of the UTM catchment and any point within the drainage system (downstream) which have high possibilities of a flash flood occurrence. The two water level monitoring stations are located on the campus of UTM based on the options shown in the figures [3].

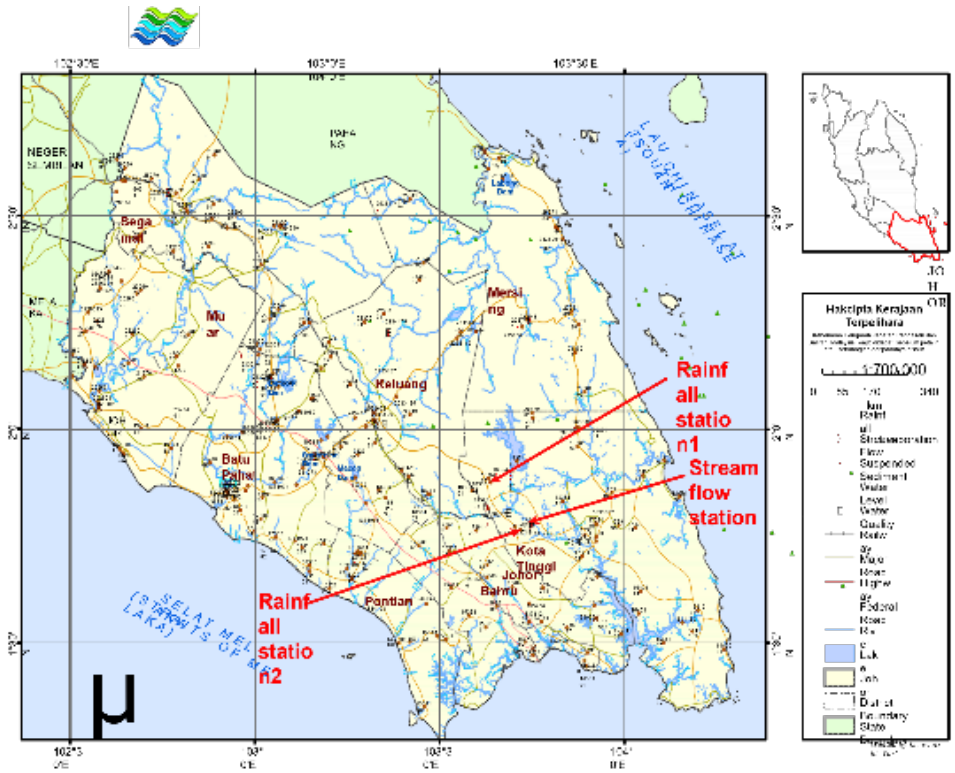

Figure 2. Hydrological Stations in JB.

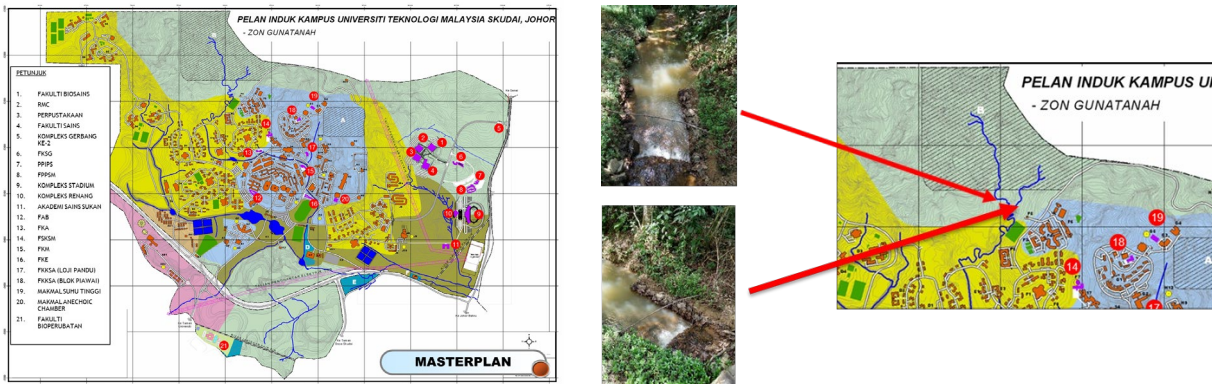

Figure 3. The location of water level sensors.

3.2 System design

The system architecture of this project, which is a real-time flood monitoring and prediction system, is depicted in Figure [4]. The system is divided into two components, physical hardware architecture and software architecture. In this system, the sensor network concept is utilized to transmit data from sensor nodes to the data logger. The sensor nodes act as a platform for data collection and measurement of the present water level and precipitation. The purpose of a data logger is to continuously monitor and record environmental factors, allowing for the measurement, documentation, analysis, and validation of conditions. Then, using a GSM modem, the data is transmitted to the cloud. Following that, the data is shown through an online server.

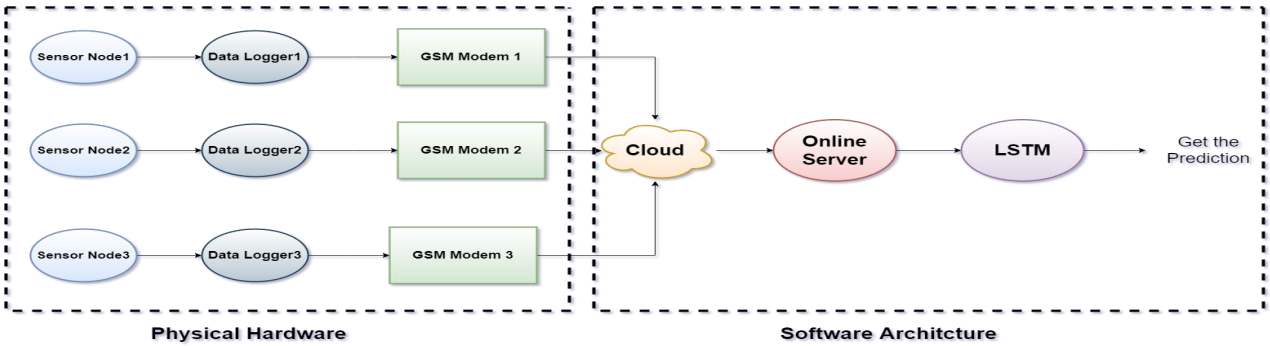

Figure 4. System architecture.

The hardware design explains the overall hardware and devices that have been used in project development based on the architecture design. In our project, we used RK400-01 Metal Tipping Bucket Rainfall Sensor (TBRS) shown in figure [5]. it depicts our sensor station installation. where the sensor is powered by a solar panel. since the rain gauge's reliance on a balanced pivot, it is critical to put tipping bucket rain gauges on a level surface. The surface should be stable and vibration-free. We proposed RKL-01 Submersible Liquid Level Transmitter (SLT) for our project, as presented in the figure [5]. The sensor operates by sensing the liquid medium's hydrostatic pressure. Hydrostatic pressure, referred to as head pressure, is the force exerted by the fluid within the vessel. SLT sensor could be installed in two liquid mediums, either in static water or dynamic water. in our case, SLT sensor is installed in the dynamic water.

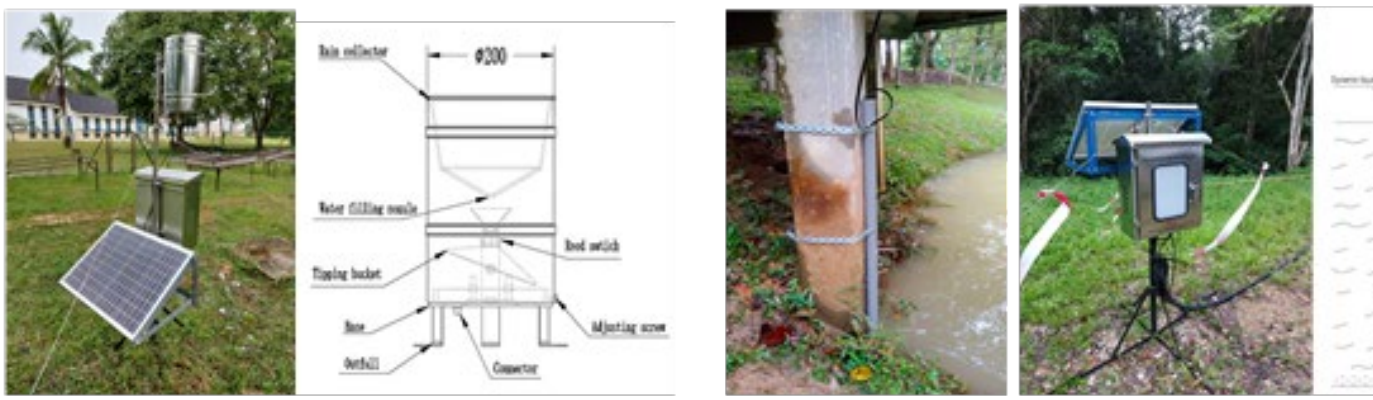

Figure 5. TBRS and SLT instillation.

This project's software development has two parts. The first part is to create a dashboard as an online server database. secondly, Develop an LSTM algorithm for flood forecasting. Figure [6] shows how the data is collected in the developed dashboard. We only collect data and save it as CSV file. The data is recorded every minute timestamp. But it was only for one month, so to do an efficient flood prediction, we will used historical data.

Figure 6. The developed online dashboard.

The reason for using LSTM is that we are working with sequential data. When we have sequential data with a timestamp for every sample, we consider this data to be time-series data. LSTM are better able to model the time domain by allowing a sequence of input vectors to be treated as a single logical input for a LSTM model. LSTM could be used for classification, regression, and generating novel output. in our case we will use LSTM for regression problem. Here in figure [7] we have our LSTM system flowchart

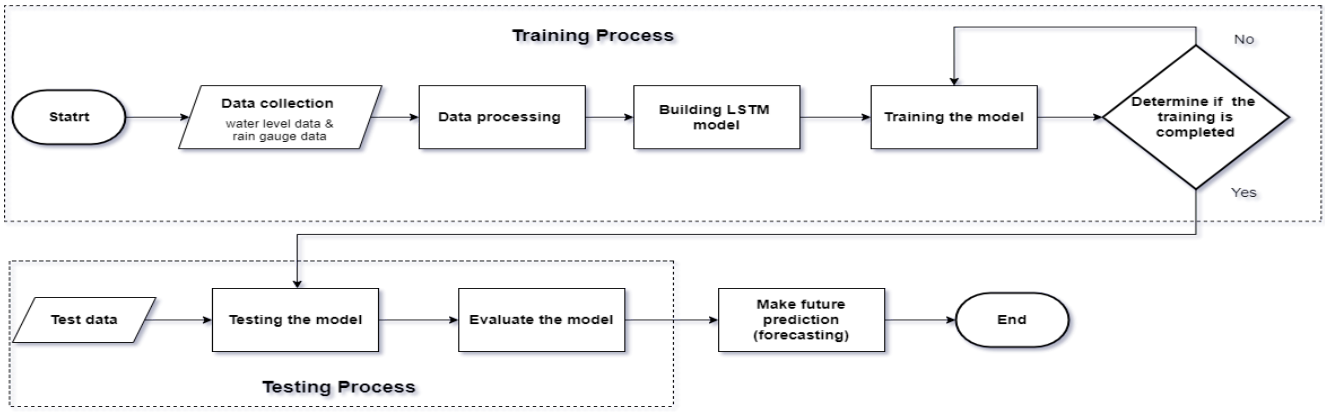

Figure 7. LSTM system flowchart.

4. Results and Discussion

This section presents the results of LSTM models for flood prediction for both historical data and the new data. Then we will choose the most accurate model to be considered.

4.1 LSTM Based on Historical data

First, we visualized the original historical data in the following figures, for streamflow figre and rainfall 1,2. we can see the maximum value of streamflow is more than 300 m/s^3 and the maximum values of Rainfall 1,2 are 73 mm and 87 mm respectively.

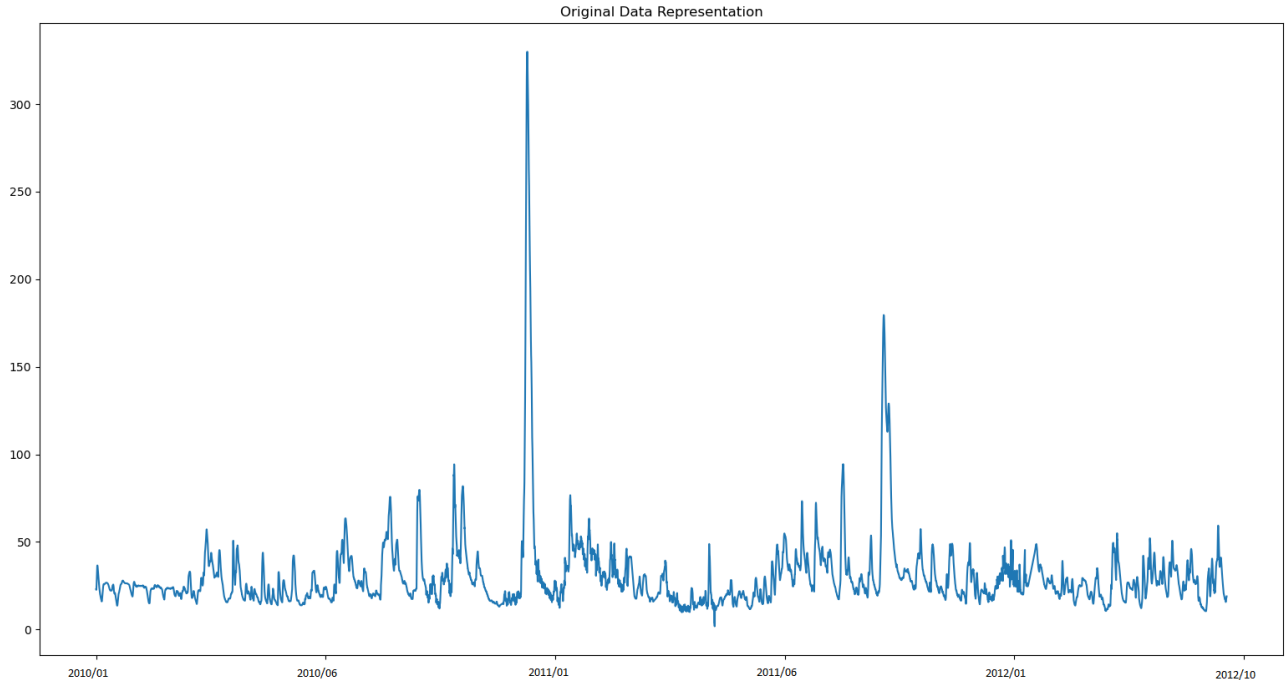

Figure 8. Streamflow data representation.

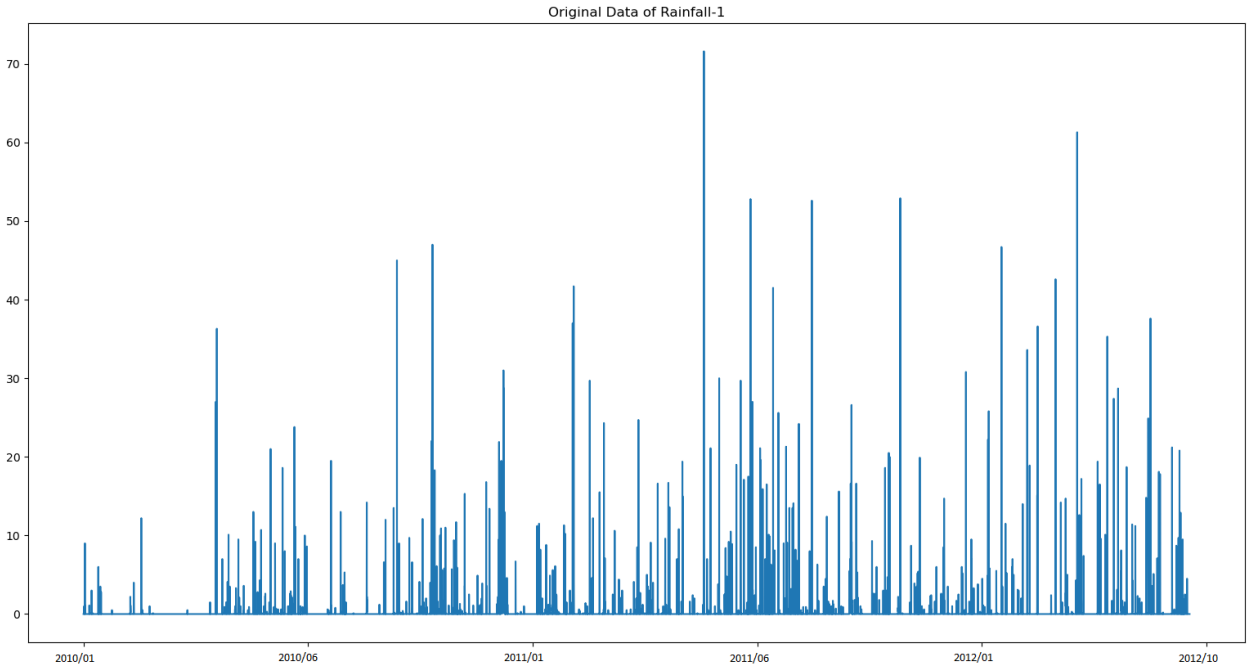

Figure 9. Rainfall-1 data representation.

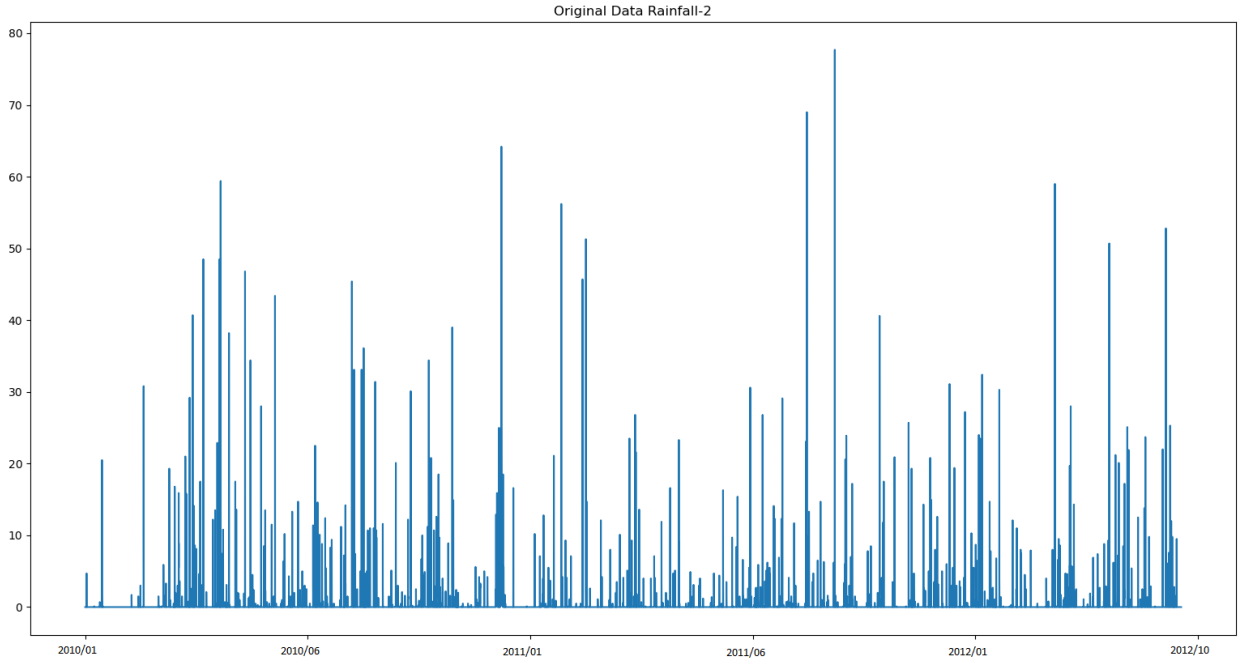

Figure 10. Rainfall-2 data representation.

The simplest model we started with is model 2x2. that means, the input is 2 rows of data, and the output is 2 data. in other words, the input is 2 hours, and we get 2 hours prediction in the output. The following illustration [9] depicts the outcome of model 2x2 and its evaluation.

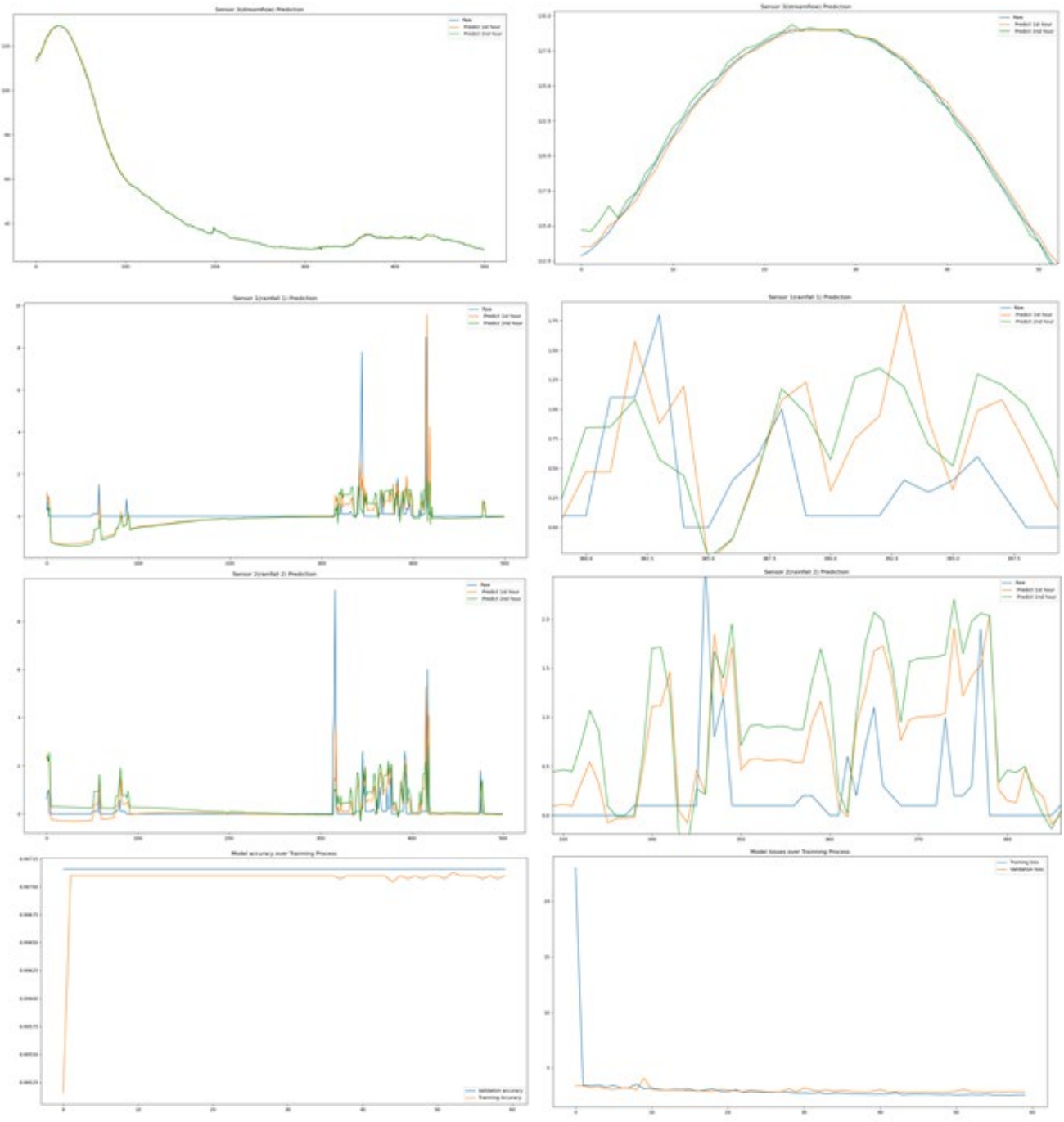

Figure 11. Model 2x2 results.

Starting with a 2x2 model, the last figure displays the predicted streamflow and rainfall 1,2. We chose 500 samples to demonstrate the outcome clearly. If we zoom in on a particular location, we can see the forecast for the first and second hour. Where the blue line, orang line and green line represent the raw data, first hour and second hour predication respectively. According to the two lines prediction, the first hour is closer to the raw line than the second. Therefore, it is more precise. Therefore, we may conclude that the more hours of forecast, the greater the mistake of the most recent hours. The evaluation of the model reveals that its accuracy exceeds 90 percent, which is quite good. In addition, the training

loss and validation loss approach zero over time. To make our model more efficient, we have to increase the predicted output so that we have enough time before the flood happen. the next schedule [] demonstrates the models that we designed with their error.

Table 1. Models results.

model 2x2	overall	1st hour	2nd hour
MAE	0.35	0.3	0.4
RMSE	1.69	1.66	1.72
MSE	2.85	2.7	3

model 2x4	overall	1st hour	2nd hour	3rd hour	4th hour
MAE	0.4925	0.4	0.46	0.51	0.6
RMSE	1.755	1.68	1.72	1.77	1.85
MSE	3.1075	2.83	2.95	3.15	3.5

model 6x2	overall	1st hour	2nd hour
MAE	0.3	0.3	0.3
RMSE	1.7	1.7	1.7
MSE	2.8	2.8	2.8

model 6x4	overall	1st hour	2nd hour	3rd hour	4th hour
MAE	0.3925	0.3	0.4	0.4	0.47
RMSE	1.7	1.6	1.7	1.7	1.8
MSE	2.9725	2.7	2.89	3	3.3

model 8x2	overall	1st hour	2nd hour
MAE	1.625	2.9	0.35
RMSE	1.685	1.67	1.7
MSE	2.86	2.8	2.92

model 8x4	overall	1st hour	2nd hour	3rd hour	4th hour
MAE	0.4175	0.3	0.4	0.46	0.51
RMSE	1.735	1.65	1.7	1.76	1.83
MSE	3.025	2.72	2.9	3.11	3.37

model 2x8	overall	1st hour	2nd hour	3rd hour	4th hour	5th hour	6th hour	7th hour	8th hour
MAE	0.775	0.6	0.8	0.8	0.8	0.8	0.8	0.8	0.8
RMSE	2.1125	1.8	2	2.1	2.2	2.2	2.2	2.2	2.2
MSE	4.625	3.3	4.2	4.5	5	5	5	5	5

model 6x8	overall	1st hour	2nd hour	3rd hour	4th hour	5th hour	6th hour	7th hour	8th hour
MAE	0.45	0.3	0.4	0.4	0.5	0.5	0.5	0.5	0.5
RMSE	1.75	1.6	1.7	1.7	1.8	1.8	1.8	1.8	1.8
MSE	3.2875	2.7	3	3.1	3.5	3.5	3.5	3.5	3.5

model 8x8	overall	1st hour	2nd hour	3rd hour	4th hour	5th hour	6th hour	7th hour	8th hour
MAE	0.43	0.34	0.35	0.4	0.47	0.47	0.47	0.47	0.47
RMSE	1.79625	1.64	1.71	1.77	1.85	1.85	1.85	1.85	1.85
MSE	3.24625	2.72	2.95	3.15	3.43	3.43	3.43	3.43	3.43

According to the tables, when we fix the input and increase the output, the error rate also rises. After reaching model 2x8, the output was fixed while the input was raised. the error decreased slightly. However, the training loss and validation loss fluctuated erratically over time. Based on this, we select the model 6x4 which has the maximum output with high accuracy and validation. In the next figure [10], the result of model 6x4 has been shown.

4.2 LSTM Based on new data (primary result)

The original new data is visualized in figure [12] for rainfall and water level1,2. we can observe that the maximum value of rainfall is 117 mm, and the maximum values of water level 1,2 are 7.2, 7.7 mm respectively. Since we only have data for one month, the majority of the precipitation data is zero. Because of this, we were unable to create an accurate rainfall forecast model. In the meanwhile, we developed a model to predict water level 1,2

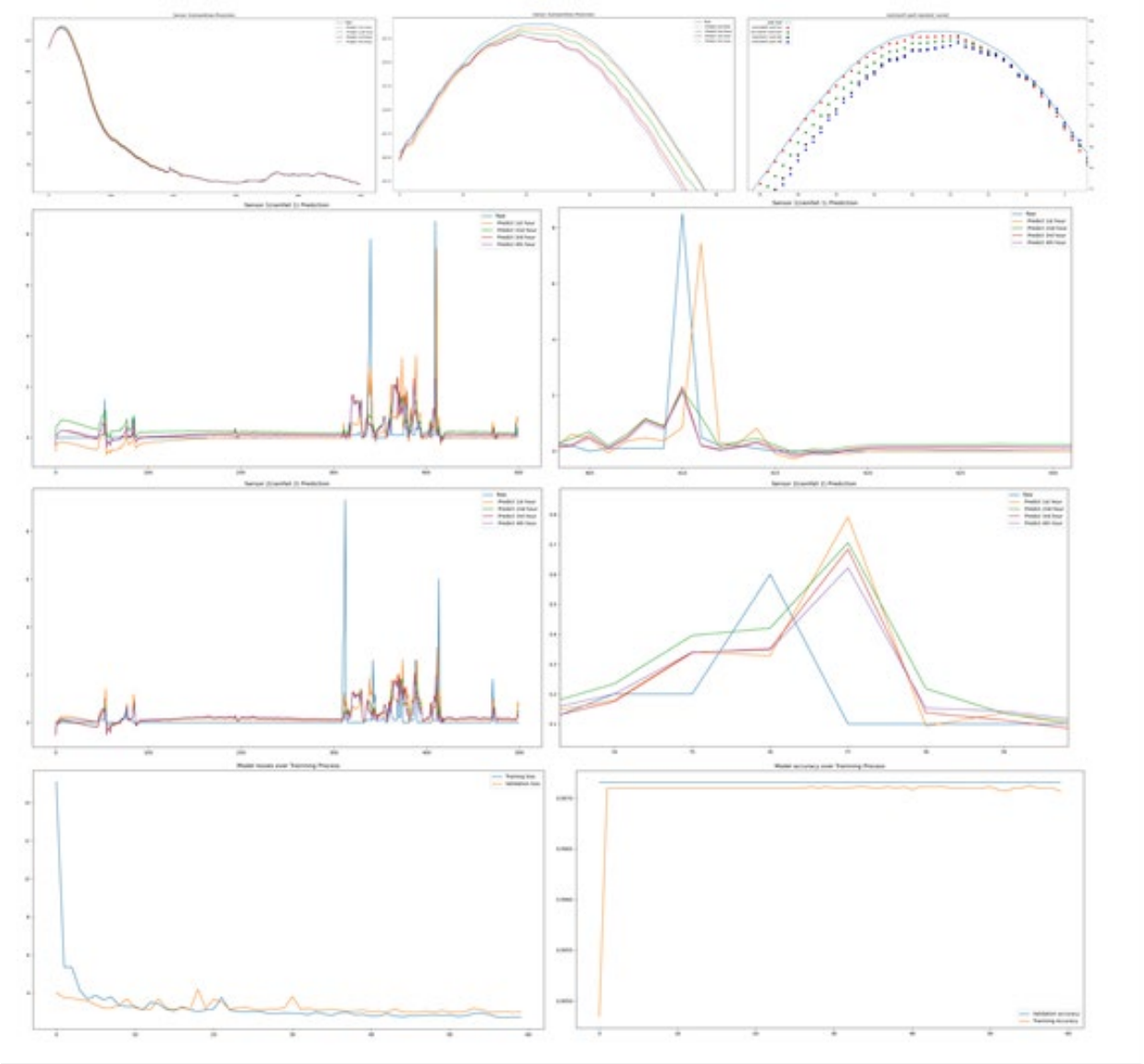

Figure 122. model 6x4 results.

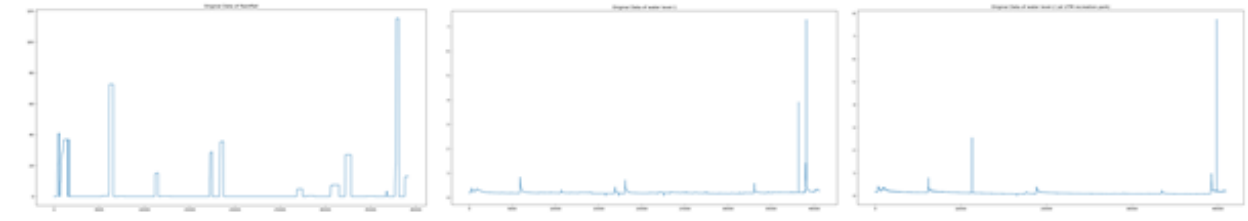

Figure 133. The original new data representation.

Using the same methodology applied to historical data, we choose the 10 x 4 model with the highest output accuracy and validation. In this scenario, the timestamp is 1 minute. hence, we enter 10 minutes and receive a forecast of 4 minutes. the following figure [12] shows the result of model 10x4 and its evaluation.

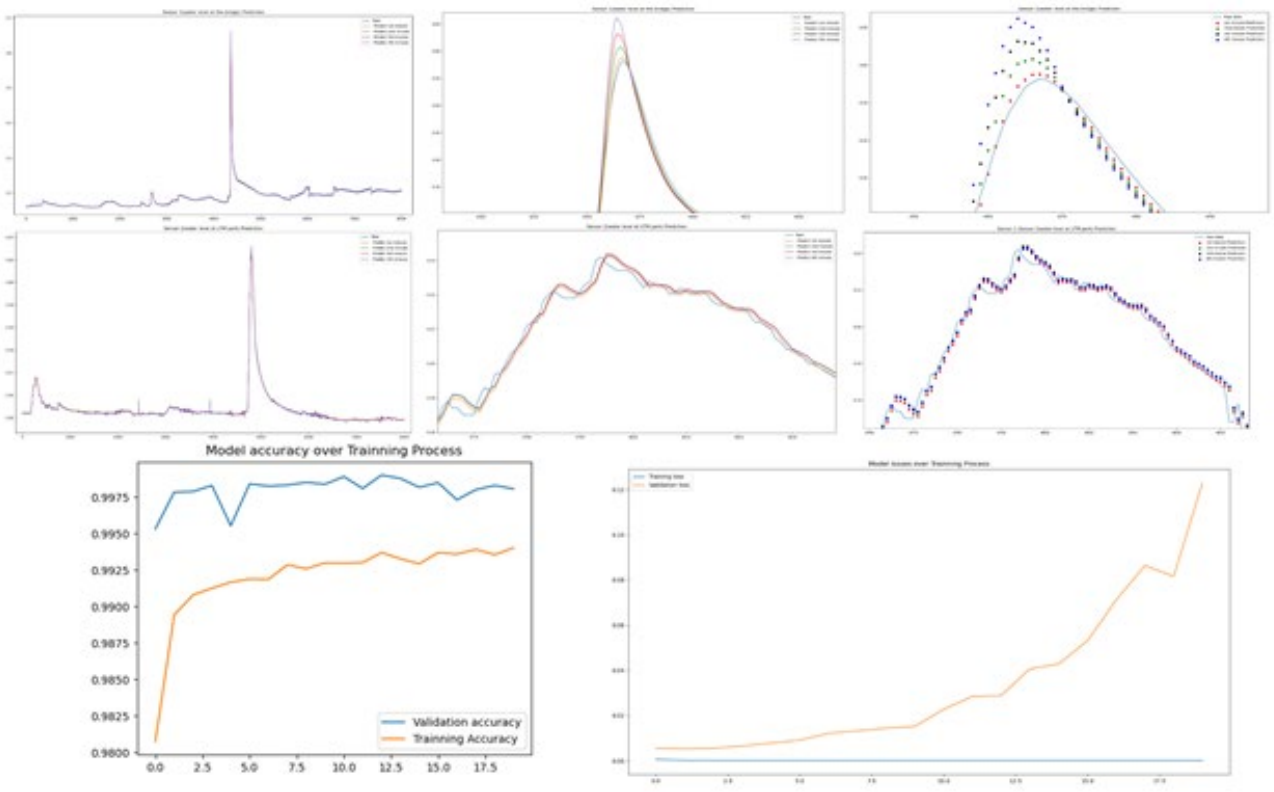

Figure 14 model 10x4 results.

The flood warning level for both water level sensors is depicted in the following graph [14]. We estimated the warning level by measuring the depth of the river at the station location. We may observe that the warning level is a significant distance from the predicted values. This indicates that the risk of flooding is quite far. This was the actual occurrence.

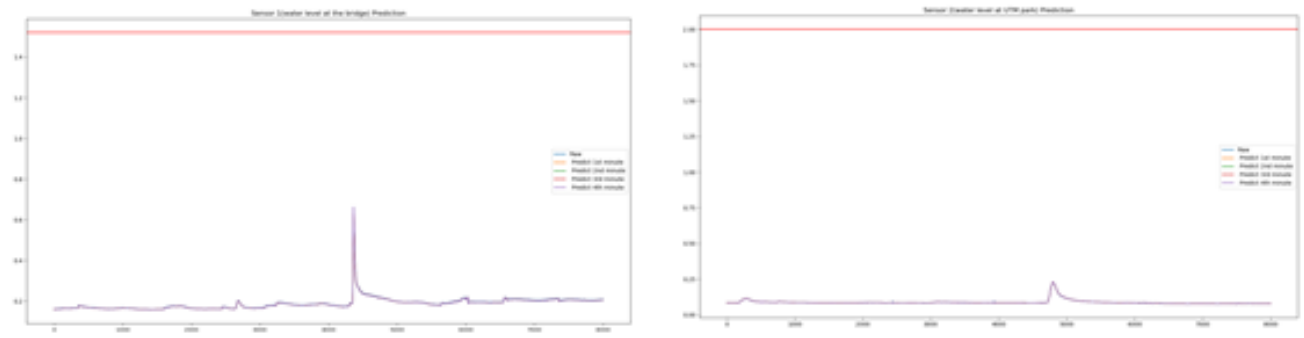

Figure 15. Flood warning level.

5. Conclusion

In this study, a monitoring system composed rain gauge and two water level sensors is developed to forecast the flood using LSTM algorithm. the results of historical data demonstrate that all the models

are suitable for prediction. model 6x4 with slightly better accuracy and validation has been chosen among other models. Results accuracy reaches above 95%. for monitoring system data, model 10x4 has been chosen as highly accurate. in future work, we can have more data for our system so that we could include the Rainfall data in prediction and forecast different lengths of time.

Acknowledgement: The authors would like to express appreciation to Universiti Teknologi Malaysia (UTM) for always providing facilities support to complete this study. We also would like to thank A2Tech Sdn. Bhd. member for their direct or indirect support in making this publication viable and possible.The authors would also like to express their gratitude to the Ministry of Education Malaysia for funding through the UTM Fundamental Research Grant [Q.J130000.2551.20H75], the Fundamental Research Grant Scheme (FRGS) (FRGS/1/2018/WAB05/UTM/02/6) [R.J130000.7851.5F032], and UTM Shine [Q.J130000.2451.09G92]

References

[1] Giovannettone J, Copenhaver T, Burns M, and Choquette S 2018 A statistical approach to mapping flood susceptibility in the Lower Connecticut River Valley Region. *Water Resources Research.* **54 (10)**:7603-7618
[2] Artan G A, Restrepo M, Asante K, and Verdin J 2002 A flood early warning system for Southern Africa. *Proc. Pecora 15 and Land Satellite Information 4th Conf*
[3] Kim G and Barros A P 2001 Quantitative flood forecasting using multisensor data and neural networks. *Journal of Hydrology.* **246 (1-4)**:45-62
[4] Campolo M, Soldati A, and Andreussi P 2003 Artificial neural network approach to flood forecasting in the River Arno. *Hydrological Sciences Journal.* **48(3)**:381-398
[5] Shah S M H, Mustaffa Z, and Yusof K W 2017 Disasters worldwide and floods in the Malaysian Region: a brief review. *Indian Journal of Science and Technology.* **10(2)**
[6] Atzori L, Iera A, and Morabito G 2010 The internet of things: A survey. *Computer networks* **54(15)**:2787-2805
[7] Mekanik F, Imteaz M, Gato-Trinidad S, and Elmahdi A 2013 Multiple regression and Artificial Neural Network for long-term rainfall forecasting using large scale climate modes. *Journal of Hydrology.* **503**:11-21
[8] Mosavi A, Ozturk P, and Chau K W Flood prediction using machine learning models: Literature review *Water*. **10(11)**:1536
[9] Gizaw M S and Gan T Y 2016 Regional flood frequency analysis using support vector regression under historical and future climate. *Journal of Hydrology*. **538**:387-398.
[10] Campolo M, Andreussi P, and Soldati A 1999 River flood forecasting with a neural network model. *Water resources research*. 35(4) 1191-1197
[11] Hochreiter S and Schmidhuber J 1997 Long Short-Term Memory. *Neural Computation*. **9(8)**: 1735-1780 doi: 10.1162/neco.1997.9.8.1735
[12] Patterson J and Gibson A 2017 Deep learning: A practitioner's approach. *O'Reilly Media, Inc*
[13] Allan M and Williams C K 2005 Harmonising chorales by probabilistic inference. *Advances in neural information processing systems.* **17**:25-32
[14] Kawakami K 2008 Supervised sequence labelling with recurrent neural networks Ph. D. thesis
[15] Rakhecha P 2007 Probable maximum precipitation for 24-h duration over an equatorial region: Part 2-Johor, Malaysia. *Atmospheric research.* **84(1)**:84-90
[16] Tan M L, Ibrahim A L, Yusop Z, Duan Z, and Ling L 2015 Impacts of land-use and climate variability on hydrological components in the Johor River basin, Malaysia. *Hydrological Sciences Journal.* **60(5)**:873-889