Artificial Neural Network Based Novel Flood Prediction Model; A Case Study in Rathnapura in Sri Lanka

KAA Chathurangi[1], LLG Chathuranga[2], and RMKT Rathnayaka[3]

[1] Department of Computing and Information Systems, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Belihuloya ,Sri Lanka
[2]Department of Computing and Information Systems, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Belihuloya ,Sri Lanka
[3]Department of Physical Sciences and Technology, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Belihuloya ,Sri Lanka

[1] Corresponding author; kaayesha@appsc.sab.ac.lk

***Abstract**— Most of the countries in the world face with natural disasters like floods, cyclones, landslides and droughts .Floods are major contributors to disruption of human lives ,economy and property damage. Also it can be strike with proper little warning or prediction. So, this study proposes a novel flood prediction model for Rathnapura district in Sri Lanka. Main reason for the flood in Rathnaura town is Kalu River. Rainfall of five meteorological stations namely Alupola station, Hapugasthenna station, Guruluwana station, Lelopitiya station and Rathnapura station are affected to the water level of the Kalu River. The methodology of this study is running under two main phases. In the first phase, K-mean clustering is used to cluster the water level of the Kalu River according to the rainfall of five meteorological stations. In the second phase, the Artificial Neural Network (ANN) model is successfully implemented for forecasting flood in Rathnapura town according to the rainfall of above mentioned five stations. Two model accuracy standards were employed. Such as mean absolute error and mean-square error. The novel ANN model gives the minimum error accuracies in both training and testing stages. Data set for this study is obtained from Department of Irrigation, Sri Lanka and it contains 1955 data. The new proposed model is useful to avoid or minimize the social and economic losses that may occur in the flood.*

***Keywords**—* **K-mean Clustering, Artificial Neural Network, Flood Prediction**

I. INTRODUCTION

Floods are kind of major natural disaster in the world. As well as Sri Lanka. In Sri Lanka major floods are associated with the two monsoon seasons. During south-west monsoon (May – Sept.) the western, southern and Sabaragamuwa provinces are vulnerable for floods. During north-east monsoon (Dec. – Feb.) the eastern, northern and north-central provinces are at risk of flooding. (A.K.W. Jayawardane et al.,2006).According to the records severe floods have occurred in the years 1913, 1940, 1957, 1967, 1978, 1989, 1992 and 2003 (Disaster Management Centre (DMC) 2005). The threats of floods have presented disruption of human lives, properties, infrastructure and economy. Problems related to flooding have greatly increased over recent decades because of population growth and the subsequent development of extensive infrastructures in close proximity to rivers. Increased frequency of extreme rainfall events, characteristics of a changing climate, can also potentially contribute to this problem (K.D.W. Nandala et al., 2009). An effective real-time flood modelling and prediction system could help mitigate the worst effects of flood disasters through the rapid dissemination of information regarding threatened areas that is simple maps of potential flood water distribution.

In Sri Lanka there are four main river basins which are vulnerable to floods. They are Kalu river basin, Kelani river basin, Gin river basin and Nilwala river basin. In each river basin there is no specific plan to face in case of a flood. Flood protection system prevailing in the country is also not at a satisfactory level. Structural measures and non-structural measures have to be improved against floods as there are series damages to the property and life. There is no Flood Hazard Maps prepared for Sri Lanka. There are no early warning speaker systems installed in our country. But people in the vulnerable area know to evacuate to a safer place by their experience (K.D.W. Nandala et al., 2009).

In this study, proposed a flood prediction model for early detection of floods focusing the Kalu River. Kalu River, the third longest river in Sri Lanka, discharges the largest amount of water into the ocean while causing floods along its route from the most upstream major town, Ratnapura, to the most downstream town, Kalutara. In Kalu River basin, floods affect rural agricultural lands as well as highly built up urban centres such as Rathnapura town.

Rathnapura is the most prominent flood zones in Sri Lanka and it is flodded recently. One of the reasons for the flood in Rathnapura is the surrounding mountains and the steep slope from Samanal hill to Rathnapura town. Most people who live near the river and lake are affected by flood. Main reason for the flood in Rathnaura town is Kalu River. The water level of the river rises according to the rainfall of the hilly area. The highest annual rainfall depth in Sri Lanka is observed in the Kālu river basin located in wetland. In some areas above the Kalu River ,annual rainfall is over 5000 mm. The damage caused by the flood is almost like every year in the main rainy season of the river basin. Kalu River basin, floods affect not only the highly constructed city centre like the town of Ratnapura but also the rural agricultural lands.

According to the Department of Meteorology in Sri Lanka, rainfall of many areas affect to flood in Rathnapura town. When considering the previous weather reports and geographic map in the Department of Meteorology they monitored that rainfall of five major meteorological areas are affected to the water level of Kalu River which may cause to flood in Rathnapura town. Such as Alupola station, Hapugasthenna station, Guruluwana station, Lelopitiya station and Rathnapura station.

This study is running under two main phases. In the first phase, K-mean clustering is used to cluster the water level of the Kalu River according to the rainfall of above mentioned five meteorological stations. In the second phase, the novel Artificial Neural Network (ANN) model is successfully implemented for forecasting flood in Rathnapura town according to the rainfall of above mentioned five stations. Two model-accuracy standards were employed. Such as mean absolute error and mean-square error. The novel ANN model gives the minimum error accuracies in both training and testing stages. Data set for this study is obtained from Department of Meteorology, Sri Lanka and it contains 1955 data. The new proposed model is useful to avoid or minimize the social and economic losses that may occur in the flood.

The rest of the paper is organized as follows. Section 2 explains about related works under this topic. Section 3 explains about the methodology of the research including technologies and techniques used in this study. Section 4 explains about experimental results. Section 5 is discussion and Section 6 ends up with conclusion.

II. RELATED WORKS

When considering about previous research publications it shows disaster management and flood prediction has already received a lot of attention in many researchers around the world. Different models and methodologies have been introduced and developed which is in related to this subject.

In 1999 Campalo et al. developed a Neural Network model to analyse and forecast the behaviour of the river Tagliamento in Italy during heavy rain periods. The model makes use of distributed rainfall information coming from several rain gauges in the mountain district and predicts the water level of the river at the section closing the mountain district. In 2000 Borga et al. conducted a study “on the use of real-time radar rainfall estimates for flood prediction in mountainous basins” for investigate the effect of systematic mean-field and range dependent radar rainfall errors on the accuracy of runoff simulation in mountainous basins. The data used in this study were collected by the C band Doppler radar located on the Monte Grande hill in northern Italy, close to Venice and Padua. In 2003 Sabhan et al. conducted a study to examine the current status of real time hydrological models used for flood forecasting and hazard mitigation and indicate how WWW-based systems can overcome some of the limitations of existing systems. In 2004 Dimitri et al. proposed a hybrid model combining M5 model tree and Artificial Neural Network for flood forecasting for the upper reach of the Huai River in China. In 2005 Martina et al. introduced fully distributed model for flood forecasting using TOPKAPI model. TOPKAPI is a physically based, fully distributed rainfall runoff model. In 2008 Todini introduced a model conditional processor to assess predictive uncertainty in flood forecasting.

In 2006 Jayawardane discussed disaster mitigation initiatives in Sri Lanka. In 2009 Nandalal introduced a hydrodynamic model to forecast floods of Kalu River in Sri Lanka. The model determine water levels along the river from Ratnapura to 79km downstream Kalutara using the HEC-RAS hydrodynamic model.

III. METHODOLOGY

This study is running under two main phases. The Department of Meteorology in Sri Lanka identified that rainfall of many areas affect to flood in Rathnapura town. According to previous flood records and geographic map meteorology department monitored rainfall of five major areas affecting floods in Rathnapura town. Such as Alupola station, Hapugasthenna station, Guruluwana station, Lelopitiya station and Rathnapura station. In the first phase, K-mean clustering is used to cluster the water level of the Kalu River according to the rainfall of five meteorological stations. In the second phase, Artificial Neural Network (ANN) model is successfully implemented for forecasting flood in Rathnapura town according to the rainfall of above mentioned five stations.

A. *Dataset*

The daily rainfall for the past ten years (From 2007 to 2017) was obtained from the Department of Meteorology, Sri Lanka for the five selected areas. In each area rainfall was obtained separately in millimetres (mm). As well as the hourly water level of Kalu River for the past ten years (From 2007 to 2017) was collected from the Department of Irrigation, Sri Lanka. They used the water indicator in the Muwagama bridge for that and the water level is measured by meters (m).The dataset used in this study contains 1955 data.

B. Algorithms

The K-means clustering is one of the popular and simplest unsupervised machine learning algorithms. It can be used to classify semi-structured or unstructured data sets. K-mean clustering accepts the number of clusters and the initial set of centroids as parameters. The distance of each item in the data set is calculated with each of the centroids of the respective cluster. The item is then assigned to the cluster with which the distance of the item is the least. The centroid of the cluster to which the item was assigned is recalculated. One of the most important and commonly used methods for grouping the items of a data set using K-Means Clustering is calculating the distance of the point from the chosen mean. This distance is usually the Euclidean Distance.

$$d_{euc}=\sum_{i=0}^{n}\sqrt{(x_i-c_i)^2}$$

Where; d_{euc}– Euclidean Distance, $x_i - i^{th}$ Point in cluster and *i* – Number of points in cluster.

Cluster centroid is the point whose coordinates corresponds to the mean of the coordinates of all the points in the cluster [H.I. Arumawadu,2005].

Artificial Neural Networks tries to capture the structure and procedure of the human brain's problem solving skill and apply them to information systems. The fundamental of the Back propagation method is to create a given function by adjusting internal weightings of input signals to compose a desired output signal. The neural network model is trained using a supervised learning method. In here potential outputs of the algorithm are already recognized and the data set used to learn the algorithm is already identified with correct results. Technically, the Back-propagation algorithm is a method for training the weights in a multilayer feed–forward neural network.

C. Training and Testing Algorithms

This study used daily rainfall data and hourly water level of Kalu River from 2007 to 2016 for training the algorithms. For the testing purposes we used data in year 2017.

D. Model Evaluation

Performance evaluation is an essential task in machine learning process. So in this study we used accuracy, Mean Absolute Error (MAE) and Mean Squared Error (MSE) for the model evaluation.

IV. RESULTS

The study is carried under the two main phases. In the first phase, cluster the water level of the Kalu River according to the rainfall of five meteorological stations as shown in Figure1 to Figure 5. X axis represents the rainfall of each station in millimetres (mm) and Y axis represents the maximum water level of the Kalu River in meters (m). According to Figure 1 to figure 5 mainly 3 clusters can be identified.

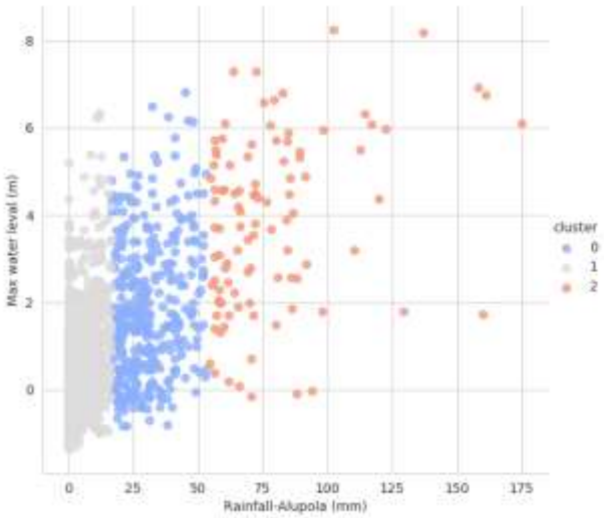

Figure 1. Clusters in Alupola Station

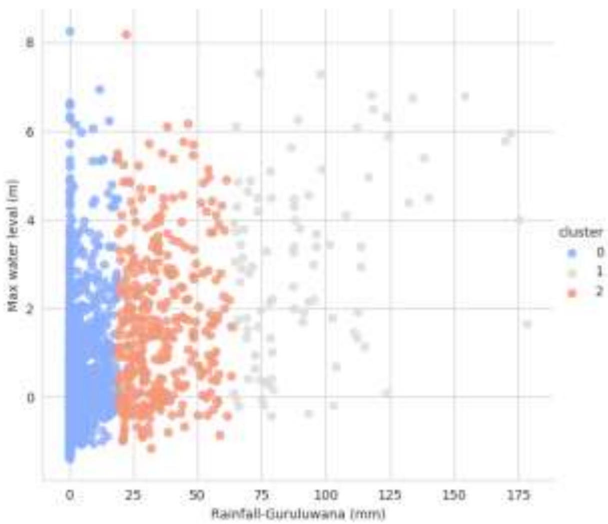

Figure 2. Clusters in Guruluwana Station

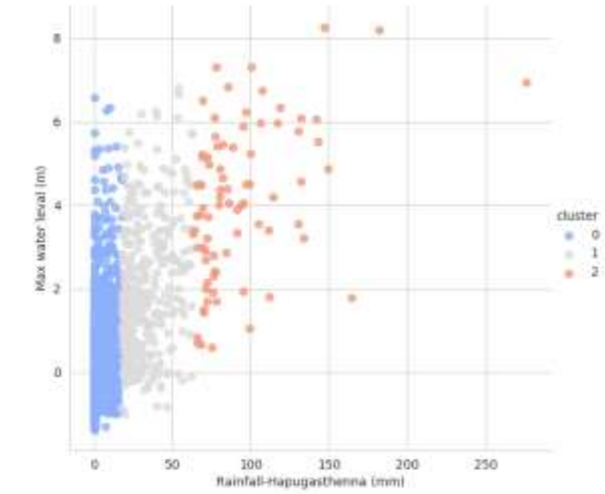

Figure 3. Clusters in Hapugasthenna Station

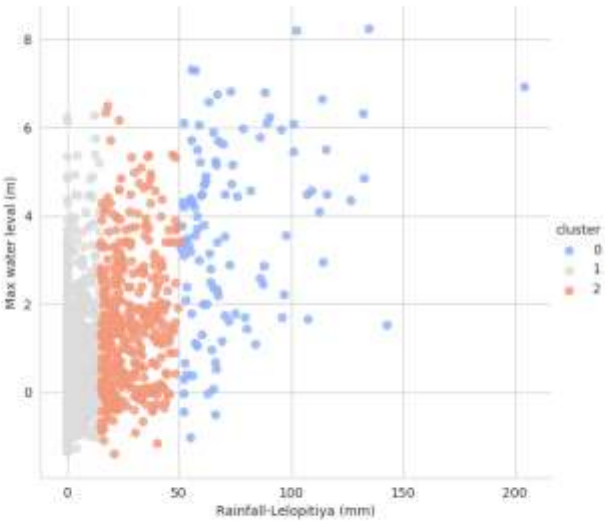

Figure 4. Clusters in Lelopitiya Station

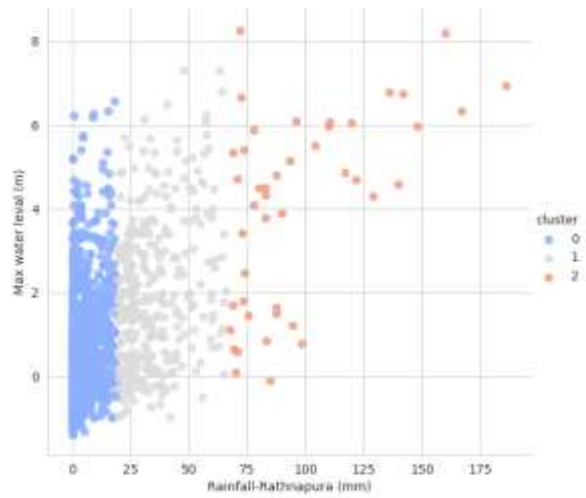

Figure 5. Clusters in Rathnapura Station

In the second phase, after a successful training a neural network model was introduced for forecast the flood in Rathnapura district according to the rainfall in five meteorological stations. It consist one input layer, three hidden layers and one output layer with three output nodes. Input layer has five input nodes and first second and third hidden layers consist 50, 50 and 45 nodes respectively as shown in Figure 6. In final ANN model an epoch is set to 60. Model accuracy for training dataset and testing dataset is 97% and 96% respectively. Mean Absolute Error (MAE) and Mean Squared Error (MSE) in this model is 2.93% and 2.07% respectively.

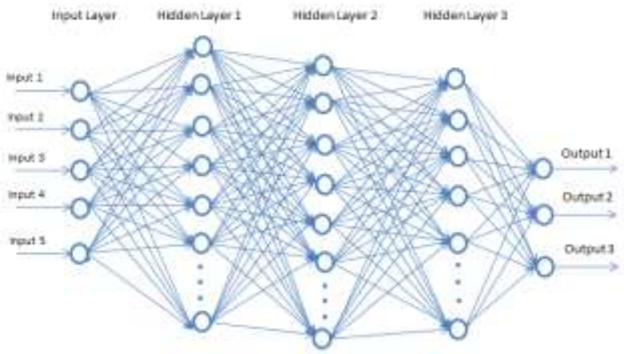

Figure 6. ANN structure for the model

V. DISCUSSION

K-mean clustering is a type of supervised learning algorithm that can be used in when you have data without defined categories or groups (unlabelled data). It will confirm what type of groups or categories exist. Also it can identify unknown groups in complex data sets. Once the algorithm has been run and the groups are defined, any new data can be easily assigned to the correct group. In this study K-mean clustering is successfully applied to cluster the rainfall in five meteorological stations that affect to the water level of Kalu River. And Back-propagation Neural Network was used to develop the flood prediction model. When comparing other machine learning algorithms ANN shows the best accuracy as shown Table 1.

Table 1. Comparison of algorithms

Algorithm	Accuracy(Training)	Accuracy(Testing)
Logistic Regression	81%	42%
Naive Bayes	85%	52%
Decision Tree	90%	61%
SVM	89%	47%
Random Forest	90%	52%

Liner Regression	96%	57%
ANN	**97%**	**96%**

VI. CONCLUSION

Floods are major contributors to disruption of human lives, economy and property damage. Also it can be strike with proper little warning or prediction. So, this study proposes a novel flood prediction model for Rathnapura district in Sri Lanka. In this study novel flood prediction model was successfully introduced. The novel model gives 97% and 96% accuracy for training dataset and testing dataset respectively. Mean Absolute Error (MAE) and Mean Squared Error (MSE) in this model is 2.93% and 2.07% respectively. The proposed model is useful to avoid or minimize the social and economic losses that may occur in the flood.

ACKNOWLEDGEMENT

Success of this research depends on the encouragement and guide that I received from the many people. I would like to thank Department of Irrigation, Sri Lanka for their contribution in data gathering for this research work.

REFERENCES

Al-Sabhan, W., Mulligan, M. and Blackburn, G. (2003). A real-time hydrological model for flood prediction using GIS and the WWW. *Computers, Environment and Urban Systems*, 27(1), pp.9-32.

Arumawadu, H., Rathnayaka, R. and Illangarathne, S. (2015). K-Means Clustering For Segment Web Search Results. *International Journal of Engineering Works*, 02(08), PP. 79-83.

Arumawadu, H., Rathnayaka, R. and Illangarathne, S. (2015). Mining Profitability of Telecommunication Customers Using K-Means Clustering. *Journal of Data Analysis and Information Processing*, 03(03), pp.63-71.

Borga, M., Anagnostou, E. and Frank, E. (2000). On the use of real-time radar rainfall estimates for flood prediction in mountainous basins. *Journal of Geophysical Research: Atmospheres*, 105(D2), pp.2269-2280.

Campolo, M., Andreussi, P. and Soldati, A. (1999). River flood forecasting with a neural network model. *Water Resources Research*, 35(4), pp.1191-1197.

Jayawardane, A.K.W. (2006). Disaster Mitigation Initiatives in Sri Lanka. *Proceedings of International Symposium on Management Systems for Disaster Prevention,*pp.09-11.

Liu, Z., Martina, M. and Todini, E. (2005). Flood forecasting using a fully distributed model: application of the TOPKAPI model to the Upper Xixian Catchment. *Hydrology and Earth System Sciences*, 9(4), pp.347-364.

Nandalal, K. (2009). Use of a hydrodynamic model to forecast floods of Kalu River in Sri Lanka. *Journal of Flood Risk Management*, 2(3), pp.151-158.

Seneviratna, D.M.K.N. and Rathnayaka, R.M.K.T., Rainfall Data Forecasting By SARIMA and BPNN Models. (2017). *IOSR Journal of Mathematics (IOSR-JM).*13(06), PP 57-63.

Solomatine, D. and Xue, Y. (2004). M5 Model Trees and Neural Networks: Application to Flood Forecasting in the Upper Reach of the Huai River in China. *Journal of Hydrologic Engineering*, 9(6), pp.491-501.

Todini, E. (2008). A model conditional processor to assess predictive uncertainty in flood forecasting. *International Journal of River Basin Management*, 6(2), pp.123-137.

BIOGRAPHY OF AUTHORS

K.A.A. Chathurangi is graduated from Department of Computing and Information Systems, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka. She is interested to do the researches in the sections of Image Processing, Data Mining and Artificial Intelligence.

L.L.G. Chathuranga is graduated from Department of Computing and Information Systems, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka. He is interested to do the researches in the sections of Image Processing, Data Mining, Machine Learning and Social Media.

R.M. Kapila Tharanga Rathnayaka is presently working as a Senior Lecturer in Statistics, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka. He highly interested to do research in the fields of Financial Mathematics, Time series Modelling and Data Mining approaches.