Advance Flood Detection and Notification System based on Sensor Technology and Machine Learning Algorithm

Mohammed Khalaf, Abir Jaafar Hussain, Dhiya Al-Jumeily, Paul Fergus, Ibrahim Olatunji Idowu
Applied Computing Research Group, School of Computing and Mathematical Sciences
Liverpool John Moores University
Byrom Street, L3 3AF, UK.
M.I.Khalaf@2014.ljmu.ac.uk , {a.hussain, d.aljumeily, P.fergus}@ljmu.ac.uk, I.O.Idowu@2009.ljmu.ac.uk

***Abstract* — Floods are common natural disasters that cause severe devastation of any country. They are commonly caused by precipitation and runoff of rivers, particularly during periods of excessively high rainy season. Due to global warming issues and extreme environmental effects, flood has become a serious problem to the extent of bringing about negative impact to the mankind and infrastructure. To date, sensor network technology has been used in many areas including water level fluctuation. However, efficient flood monitoring and real time notification system still a crucial part because Information Technology enabled applications have not been employed in this sector in a broad way. This research presents a description of an alert generating system for flood detection with a focus on determining the current water level using sensors technology. The system then provides notification message about water level sensitivity via Global Communication and Mobile System modem to particular authorise person. Besides the Short Message Service, the system instantaneously uploads and broadcast information through web base public network. Machine-learning algorithms were conducted to perform the classification process. Four experiments were carried out to classify flood data from normal and at risk condition in which 99.5% classification accuracy was achieved using Random Forest algorithm. Classification using Bagging, Decision Tree and HyperPipes algorithms achieved accuracy of 97.7 %, 94.6% and 89.8 %, respectively.**

***Keywords* - Flood detection system; Sensor network; Short message service; Global communication and mobile system; Real-time data; Machine learning algorithm**

I. Introduction

Flood is considered as one of the most common natural phenomenon in the world, affecting human lives and causing severe economic damage to goods and properties. Preventing floods before actual happen can provide sufficient time for inhabitants evacuate in the nearby areas. Forecasting and warning system could have potential impact to mitigate the severe of flood affects. With this context, there is an important need for sharing experience and collaboration among countries in order to mitigate the effects of floods before they proceed to the critical condition.

This ageing natural disaster cannot be escaped but proper managing and pre-alarming system can mitigate its severity. In most of the developing countries, flood monitoring cell under meteorological department are not properly equipped with smart and scalable flood alarming system. As a result, people from flood affected prone areas are suffering the consequences of flood every year. To determine the river water levels and detect flood, IT applications can provide valuable solutions. Moreover, notifying people through proper and efficient way in real time is another inevitable task to defy flood devastations.

To date, sensor technology is widely utilised for detecting water level variations in different flood prone zone. However, wireless communication (such as GSM, SMS etc.), microcontroller, web based applications are yet to be enhanced for detecting and pre-notifying about river inundation to the people. As a requirement, sensors are utilised to measure the level of water either in dangerous zone or normal condition. Besides these, efficient and cost effective flood alarming systems are not very common in many developing countries till now. Therefore, the unavailability and ineffective utilization of alarming system have caused severe damages recently in Philippine [1], Thailand [2] and Brazil [3] that ultimately brings losses of lives and properties. Consequently, a suitable alert tool can provide lot of benefits and offer continuous alert information before the situation becomes critical and destroy lives and properties.

There are several equipment that can be combined together to send notification using water level sensor to the emergency management authority. Water level sensor, a Programmable Logic Controller (PIC), a Global Communication and Mobile System (GSM) modem, a power supply and Subscriber Identity Module (SIM) were used in this research. The main idea behind utilising GSM technology is to trigger the water level severity based on water sensor in real time via SMS service.

The remainder of this paper is organised as follows. Section II shows the system design and description of the proposed system. Hardware experiment result is illustrated in section III whereas the methodology is presented in section IV. Section V demonstrates the results. Finally, conclusion shows in section VI.

II. System Design and Description

An effective early flood warning and forecasting system can be valuable to the community as it act as a protective action for the victims in terms of preventing loss of civilian casualties and infrastructures. SMS is a proper alert communication tool that can disseminate the information to flood victims within particular area. Figure 1 shows the complete process flow for hardware and software development of the proposed framework architecture. Three water level sensors have been represented in order to provide real-time information to the flood control centre for processing purposes. Each sensor has a special function. The first sensor detects the normal level of water, while the second sensor detects the above normal, and the third sensor detects the dangerous condition.

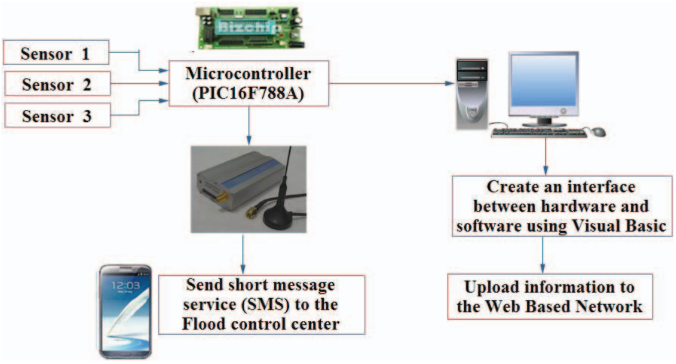

Figure 1. System arichtiture

Data collected from various sensors transmitted to the main control unitt (microcontroller PIC). PIC receives various signals from three sensors and converts them from analog to digital using a special convertor. This system is depending on microcontroller, hence improvement in software is considered essentials. The experimental setup can be controlled and monitored from anywhere covered by GSM service by exchanging SMS with the server.

A. Signal Processing

The three different sensors will automatically produce electric pulses once they detect any change in water level. PLC will modify and compile electric pulses into digitisation process before proceeding it to the flood warning services. PIC device will be configured to send instant SMS as output whenever it receives an input from the sensor using GSM modem to inform the flood warning services about water condition. This research intends to improve hydrology forecasting to save lives and economic system by alerting residents who live nearby the potential flood area. The components are connected together to the main system as illustrated in Figure 2.

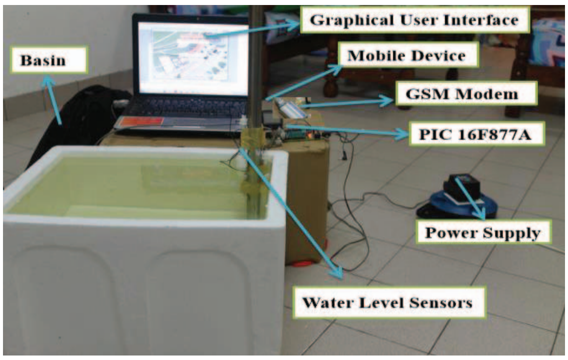

Figure 2. Main board of flood detection system

III. Hardware Experiment Results

The three water sensors are placed at different levels and locations and can be used to predict flood. Text message services was selected as the long range communication protocol. SMS alert will send every 30 minutes in order to inform flood management centre about current conditions. The activation time between second and third sensor is depending on water level condition. There are three different data input will be received from the various sensors for warning or non-warning messages as illustrated in Figures 3.

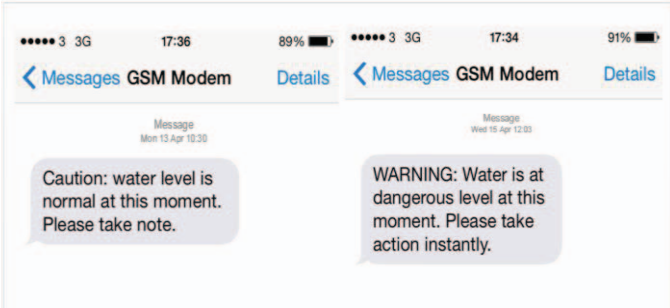

Figure 3. Warning message by SMS

A. Testing Results with Visual Basic

It is essential to use different kinds of tools for predicting the environmental conditions in particular with flood phenomena. In the monitoring screen, the PLC will receive raw data from sensors then compiled, analysis and display the current status in the GUI. If the sensor detects normal level, the warning alert appears automatically on the main interface "the water is normal" as demonstrated in Figure 4. If the water level exceeds normal level, the warning alert appears automatically "water Level is Normal". Finally, if the sensor detects water exceeding dangerous level, the warning alert message appears automatically on the main interface "water Level is Dangerous". Therefore, the flood monitoring services can monitor any changes of water level and give immediate response once water level reaches dangerous zone.

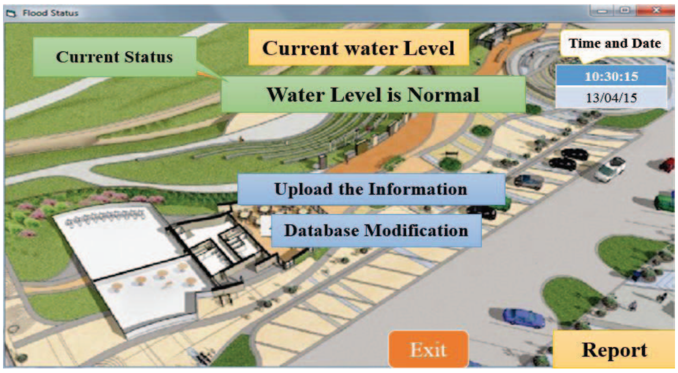

Figure 4. GUL display for water level

B. Website Interface

Website is considered such a powerful tools to disseminate information. This is used to make the vast majority of inhabitants well informed and provide the latest news about flood current conditions. This real time web base portal announcement results is considered a significant notification for flood prone zones that ultimately allow the public to have ample time to move to safer zones. Figures 5 illustrates the normal, above normal and dangerous level.

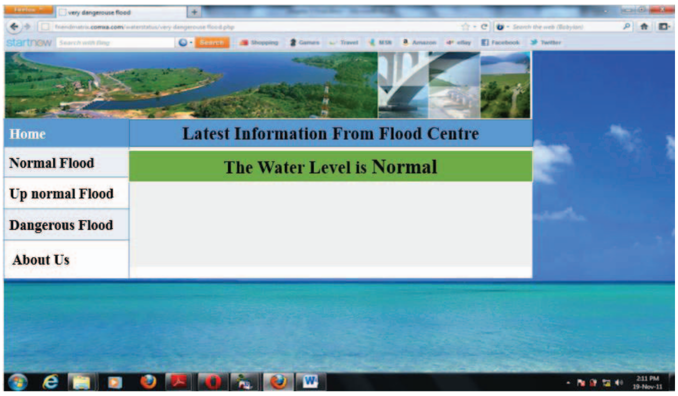

Figure 5. Waring message on web based network

IV. Methodology

There are two major parts in our experimental approach. The first part of the experiment depend on hardware components in order to collected raw signal using sensor technology, PLC and GSM. The second part of the experiment is based on using machine learning algorithm to analyse the level of flood data to quantify if the level of water is normal or dangerous condition.

In order to collect datasets for classifying flood data with normal and up-normal condition, raw data has been extracted from environment agency website that is responsible for flood warnings, river levels and flood risk in the United Kingdom [4]. Table 1 summarised the most important attributes that usually flood monitoring services concentrate on to check flood disaster. In these experiments, we used 1000 flood records that were collected within 5-year period with 623 under risk conditions and 377 non-risky condition records.

TABLE I. CHARACTRISTICS OF FLOOD

No	Types of Attributes
1	Weather
2	Tides
3	Season
4	Months
5	Measure
6	Value
7	Min flood depth (Metres)
8	Max Flood depth (Metres)
9	Southwest monsoon
10	Northeast monsoon

A. *Confusion Matrix Evaluation*

To examine the proposed machine learning techniques for the classification of flood data, contingency table was utilised (known as confusion matrix and matching matrix) to analyse the datasets. In this study, four machine-learning algorithms were implemented. These algorithms are Random Forest [5], Bagging [6], Decision Tree [6], and HyperPipes [7] .

The True Positive (TP) and True Negative (TN) illustrate the most correct classification of positive instance and negative instance, respectively. False Positives (FP) shows negative symbols while, False Negatives (FN) demonstrates the positive instance.

There are number of measurements that can be used to evaluate the model performance in terms of accuracy. The accuracy can be calculated as follows:

$$AC = \frac{TP+TN}{(TP+FP)+(FN+TN)} \quad (1)$$

While equation (2) is used to evaluate the proportion of positive instances that were correctly classified.

$$TP = \frac{TN}{(FN+TN)} \quad (2)$$

The proportion of negative instances that were classified incorrectly can be determined as:

$$FP = \frac{TN}{(TP+FP)} \quad (3)$$

In order to classify the TN that were classified correctly, equation (4) were carried out to calculate the datasets:

$$TN = \frac{TP}{(TP+FP)} \quad (4)$$

The equation (2) was conducted to evaluate the proportion of positive instances that were incorrectly classified.

$$FN = \frac{FN}{(FN+TN)} \quad (5)$$

B. *Data Collection*

As mentioned previously, the original datasets that are used in our experiments were gathered within five-year period from the environment agency website, UK. Each sample consists of 10 parameters deemed vital factors for predicating flood catastrophe as illustrated in Table 1. We applied different machine learning techniques on the datasets to evaluate the complete classification in terms of strengths and weaknesses; using contingency table (i.e., accuracy, Recall, aROC etc.).

V. RESULTS

In this section, we present the classification outcomes using four techniques.

As mentioned before, four algorithms with 10 fold-cross validation have been used to classify flood data as shown in Table 2. It is necessary to note that four experiments was conducted with imbalanced datasets (623 at risk: 377 no risk). The first correct classified Instances obtained 89.8% using HyperPipes algorithm, whereas incorrect classified Instances is

10.2%. Secondly, Decision Tree algorithm presented better outcomes than HyperPipes algorithm. The correct classified showed 94.6%, while, incorrect classified Instances were 5.4%. Furthermore, 97.7% of the correct classified Instances were achieved by applying Bagging algorithm, whereas 2.3% incorrect classified. Finally, Random Forest achieved the best results among other classifications with 99.5 % for correct classified Instances, while, incorrect classified Instances indicated only 0.5%.

TABLE II. CLASSIFIER OUTPUT FOR CLASSIFICATIONS

Parameters	Methods			
	Random Forest	Bagging	Decision Tree	HyperPipes
Correctly Classified Instances	99.5 %	97.7 %	94.6%	89.8%
Incorrectly Classified Instances	0.5%	2.3%	5.4%	10.2%
Kappa statistic	0.9894	0.951	0.8865	0.7939
Mean absolute error	0.0233	0.0417	0.1203	0.4457
Root mean squared error	0.0902	0.1386	0.1961	0.4475
Relative absolute error	4.9494 %	8.872 %	25.6159 %	94.8586 %
Root relative squared error	18.6146 %	28.5952%	40.4556 %	92.3362 %
Total Number of Instances	1000	1000	1000	1000

Tables 3 shows the accuracy by class of classifications. In order to maximise information advantage from the datasets, we used a number of applications to identify the proper techniques that produce high performance and accuracy.

TABLE III. ACCURACY OF CLASSIFICATIONS

Methods	Accuracy By Class				
	True Positive / Sensitivity	True Negative/ Specificity	Recall	F-Measure	ROC Area
Random Forest	0.997	0.006	0.997	0.993	0.998
Bagging	0.968	0.018	0.968	0.969	0.996
Decision Tree	0.96	0.063	0.96	0.931	0.989
Hyper Pipes	1	0.164	1	0.881	0.982

The Random Forest algorithm achieved more effective and tends to yield better outcomes in comparison to other classifications because of the possibility of solving problems that have some complexity and easy to use. The algorithm can deal with supervised learning algorithm technique which can ensemble classifier utilising large number of decision tree models in order to prevent overfitting problem and producing easy to set parameters. The main idea behind using this method is to reduce the error until the Random Forest learns the complete training dataset correctly and efficiently. Our motivation behind utilising Random Forest algorithm is that it can be modified during the training to accommodate for more data. Furthermore, it can classify a large number of new datasets more rapidly. In this case, our extensive research indicated that the performance of Random Forest compared with other classifications is considerably more beneficial as it offers more accurate outcomes.

VI. CONCLUSION

Flood detection system has been designed for immediate notification to the local authorities. It determined the current water level using sensor network, which provides notification via SMS and web base public network through GSM modem. SMS and web base public network are valuable alert communication tools that can distribute the information to the floods victims within particular area.

Four machine-learning algorithms were utilised to classify flood data. Random Forest algorithm provides a significant improvement in terms of accuracy over other classification with 99.5 % accuracy in comparison to 97.7% using Bagging algorithm. Moreover, Decision Tree achieved 94.6% accuracy, which is slightly lower than Bagging. While, HyperPipes algorithm obtained 89.8%, which is considered the lowest accuracy percentage acquired among other classifiers.

However, this project can be improved to achieve more advance and better application in the next stage of research. For future improvement, this system can be upgraded by adding GPS module to track the equipment that located in various field such as riverbank or low-lying area. Finally, Time series architecture can be implemented for machine learning algorithm in order to enhance the outcomes of project.

REFERENCES

[1] J. Aerts, W. W. Botzen, K. Emanuel, N. Lin, H. de Moel, and E. O. Michel-Kerjan, "Evaluating flood resilience strategies for coastal megacities," Science, vol. 344, pp. 473-475, 2014.

[2] S. Fakhruddin and Y. Chivakidakarn, "A case study for early warning and disaster management in Thailand," International Journal of Disaster Risk Reduction, vol. 9, pp. 159-180, 2014.

[3] L. C. Degrossi, J. Albuquerque, M. C. Fava, and E. M. Mendiondo, "Flood Citizen Observatory: a crowdsourcing-based approach for flood risk management in Brazil," in 26th Int. Conf. on Software Engineering and Knowledge Engineering, 2014, pp. 1-6.

[4] Environment Agency, "Environment Agency Offices datasearch" [Online].Available:http://maps.environmentagency.gov.uk/wiyby/dataSearchController?lang=_e&textonly=off&topic=offices".[Accessed:01/07/2015].

[5] S. E. Sesnie, B. Finegan, P. E. Gessler, S. Thessler, Z. R. Bendana, and A. M. Smith, "The multispectral separability of Costa Rican rainforest types with support vector machines and Random Forest decision trees," International Journal of Remote Sensing, vol. 31, pp. 2885-2909, 2010.

[6] M. C. Tu, D. Shin, and D. Shin, "A comparative study of medical data classification methods based on decision tree and bagging algorithms," in Dependable, Autonomic and Secure Computing, 2009. DASC'09. Eighth IEEE International Conference on, 2009, pp. 183-187.

[7] Z. A. Deeb, T. Devine, and Z. Geng, "Randomized Decimation HyperPipes," ed: Citeseer, 2010.