IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 19, Issue 5, Ver. I (Sep.- Oct. 2017), PP 11-13
www.iosrjournals.org

Enhanced Framework for Long-Lead Upstream Flood Analysis

*Ahmed Babalaji Ndanusa[1], Zulkhairi B. Md Dahalin[2], Azman Ta'a[3]
[1,2,3]*(Dept. of Information Technology, College of Arts and sciences, Universiti Utara Malaysia, Malaysia).*
Corresponding Author: Ahmed Babalaji Ndanusa

Abstract: *Current flood mitigation approaches consist of responding aptly to flooding events with either upstream or downstream measures. Upstream factors such as precipitation and other hydrological factors are the major triggers of flooding events within our environments. In order to address this issue, a conceptual framework was introduced in this paper to identify various spatio-temporal data required for an enhanced flooding analysis, as well as the appropriate processes needed for analyzing any potential flooding events for apt decision-making. In identifying the importance of these contributing factors, the approaches can be placed within a coherent framework as well as highlighting the importance of every approach involved. The limitations faced by the current approaches will also be highlighted.*

Keywords: *Flood Analysis, Long-lead flood Analysis, Spatio-temporal Data, Upstream Flooding.*

Date of Submission: 16-08-2017 Date of acceptance: 08-09-2017

I. Introduction

Flooding event is identified to be one of the most disastrous natural events affecting both lives and properties of individuals within the affected area(s) [1]. Unfortunately, the frequency of these natural disasters induced by flooding will continue to increase as a result of the current global warming [2]. Consequently, a suitable means of identifying and obtaining a reliable analytical result needs to be proposed in order to mitigate the overwhelming havoc caused by floods [3].

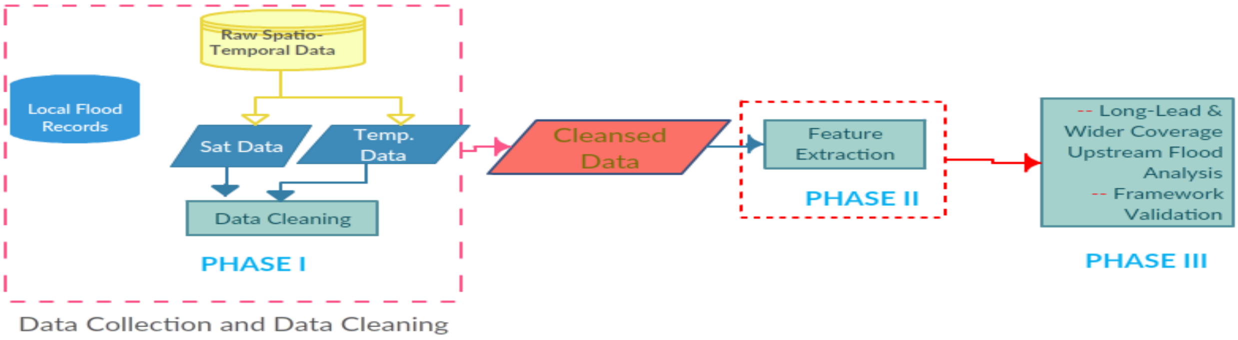
Currently, mitigating approaches have been focused on non-structural means such as predictive frameworks instead of the traditional physical means which, when adopted, will not only mitigate the impacts on lives of individuals, but will also reduce the economic losses suffered. Interestingly, an accurate long-lead upstream flood approach will ensure a major breakthrough in mitigating the impacts of disasters associated with flood. This paper will propose a solution that can accurately perform a long-lead prediction using big data analytical approaches which will address the existing gaps in some of the current related works utilized to perform the same task. Thus, this paper will be segmented into threefolds. Firstly, the paper discusses and presents the pros and cons in some of the related works, Section two will propose a conceptual framework required for an efficient upstream flood situational analysis in a long-lead, while Section three will discuss the contribution of this study.

II. Related Works

As earlier stated, the adoption of a framework for a reliable upstream flood situation recognition has been identified to play a vital role in mitigating environmental impacts instigated by flood [4]. Therefore, this section will meticulously review some frameworks and techniques adopted in upstream flood prediction and analysis which are highly needed by metrological department and most specifically, disaster monitoring and reporting agencies whose aim is to find a more versatile model and computational assessment tools with ideal data, cost efficiency, and within a reduced time frame to achieve an accurate result. Over the past years, upstream flood situation analysis has recorded a tremendous success as well as failure due to the inability of some studies to adequately implement some pre-processing techniques within the adopted methods, which would have greatly enhanced the effectiveness of the approach [5].

The use of NNARX structure for upstream flood prediction for a time frame of 10 hours lead-time utilizing Gradient Descent as training algorithm was adopted to obtain a model by segmenting data sets into training, validation and testing data [6]. With a historical records acquired for the period of ten days totaling 1463 were used for training sample. Meanwhile, for model validation, 2000 records were acquired for a period of 13 days while 4000 record sets which were acquired for a period of 28 days were utilized as testing sample. The model was able to perform a 10-hour prediction of water level, but the prediction underestimated the water level. Additionally, a major milestone was attained in [7] by adopting a predictive model based on Best Fit Probability Distribution to evaluate the volume of rainfall and runoff using one of the three Dams. However, with an enhanced spatio-temporal data from satellite, the scope will be amplified to encompass the whole study area including the three dams present within the area as proposed in [7]**.** Despite the successes recorded in these various instances of prediction, these approaches also present a number of drawbacks, especially in terms of the lead-time required for flood analysis which were between 5 to 10 hours, as some proactive measures, such as provisional bridges as well as alternative settlements plans for the evacuees cannot be adequately implemented within the periods generated by these studies. More so, the standard lead-time of an upstream flooding ranges between five to fifteen days [8]. Therefore, this study will be adapting the framework in Figure 1 to efficiently manage the data sets from the collection to framework validation phases. Primarily, as a result of the importance attached to data cleaning, this study will adopt geometric and atmospheric corrections to clean the raw data which were not performed in the existing studies. Additionally, to clean the spatio-temporal data before analyzing a long-lead flood situation as conducted study conducted in Iowa which utilized atmospheric and climate data alongside Decision Tree algorithm to perform a long-lead prediction of 9 days with an accuracy level 87% [9]. Therefore, this paper will propose the following framework aimed at presenting a complete approach that will address the various issues identified in the existing studies.

III. Porposed Framework



The Figure above is segmented into three tasks which represent three processes as thus:

PHASE I:

Data Collection/Cleaning: The collection of reliable and qualitative sets of satellite images as well as temporal data from a large and diverse range of sources is enormously time consuming. For the purpose of implanting the proposed framework, data sets that will be collected (Table 1) will be in a wide range of formats from paper documents and maps to other digital formats. After the collection of these data, the subsequent task in this phase will encompass the cleaning of the raw data which is aimed at ensuring and ascertaining a reduced level of error in the data sets needed for the long-lead upstream flood analysis.

The following are the required sets of data needed for the study:

Table 1

S/N	Data Set	Purpose
1	Satellite image (Raster Data)	To extract the following features: Stream accumulation Flow direction of water Flow length Water shade Basins Slope of the areas
2	Geological map	To give a detailed description on the soil type for the study area. This will determine the rate of water being absorbed by the soil when there's a heavy rainfall.
3	Daily Discharge Data of water from any major water bodies in the study area	To determine the rate of discharge of water from the water bodies which could also contribute to flood when flooded due to rainfall
4	Water Level Fluctuation	Shows the water fluctuation of level in the dam
5	Precipitation Data	Rate and volume of rainfall within the study area

6	Flood records in the study area	Records of the past flooding events within the study area. This consists of the date, and the communities affected by the flooding events
7	LandSat Data	Land Use Cover Land features
8	Shape file (Administrative)	Indicating the attributes of a map with localities and regions.
9	Shape files (water bodies)	To be used to identify the presence of water in the study area

PHASE II

Feature extraction: A cleaned spatio-temporal data sets remain unusable without a proper application of a classification algorithm on the imagery. Therefore, the tasks involved in this segment is the application of unsupervised classification of the cleaned data in order to derive useful features representing the study area for a proper interpretive insight needed for a long-lead flood analysis. This will be achieved by measuring the heterogenous patterns of digital elevation models (DEMs), slope, geology, vegetation, and hydrological features in order to determine various regions prone to flood vulnerability and the corresponding level of vulnerability when induced by a relatively high volume of precipitation [11, 12].

PHASE III

Long-lead: As earlier mentioned, the standard lead-time for long-lead predictions is between five (5) to fifteen (15) days; nevertheless, some frameworks perform a longer lead-time by utilizing between medium and long-lead weather situation recognition. Additionally, basins within some regions respond slowly, thereby aiding the possibility to obtain a long-lead prediction without much computational efforts.

As this proposed framework will use voluminous and heterogenous sets of data capable of representing many features inducing floods. These features represent the basic concept of Big Data [10], while broadly, the objective of this paper is the adoption Big Data Analytics (BDA) which will aid to find a suitable environment and astute means in utilizing the extracted factors capable of causing flooding events within the study area.

Furthermore, the frequency at which a prediction is performed will be determined by the need for such situation. Almost half of the framework surveyed performed flood situation recognition only when the need arises at the moment(s) when upstream flooding events are expected. However, disaster monitoring agencies perform flood situational recognition during a stable weather to provide useful information needed by emergency responders, such as hydropower generator and recreationalists in order to take proactive measures aimed at mitigating the potential impacts of these flooding events. Therefore, the tasks involved in this phase will serve in performing an accurate long-lead prediction by observing the trend of the precipitation data the Inducible Flood Precipitation Values (IFPV) as well as assessing the extent to which the lead-time can attain for the study area while analyzing the corresponding severity of the upstream flood.

And finally, the validation of the framework will be done by comparing various levels of vulnerability with the frequency of flooding events experienced by the corresponding area as detailed in the local flooding events data set.

IV. Conclusion

In an upstream flood situational recognition, the utilization of voluminous spatio-temporal data to extract the most useful information poses a great limitation for hydrological researchers as well as data scientists. This paper aims at harnessing the potentials that dwell in Big Data Analytics(BDA) to design a long-lead framework capable of recognizing an upstream flood situation accurately. The resultant framework illustrated how time dependent spatio-temporal data within a geospatial environment can be effectively explored visually using suitable BDA framework for pre-processing spatio-temporal data. This framework includes direct manipulation tools for cleansing, extracting, and mapping elements needed for a timely recognition of upstream flood analysis required for decision-making.

References

[1] Barredo, J. I., De Roo, A., & Lavalle, C. (2007). Flood risk mapping at European scale. *Water Science and Technology*, *56*(4), 11-17.

[2] King, D. A. (2004). Climate change science: adapt, mitigate, or ignore?. *Science*, *303*(5655), 176-177.

[3] Sang, J., Allen, P., Dunbar, J., & Hanson, G. (2014). Development of semi-physically based model to predict erosion rate of kaolinite clay under different moisture content. *Canadian Geotechnical Journal*, *52*(5), 577-586.

[4] Wanders, N., Karssenberg, D., Roo, A. D., De Jong, S. M., & Bierkens, M. F. P. (2014). The suitability of remotely sensed soil moisture for improving operational flood forecasting. *Hydrology and Earth System Sciences*, *18*(6), 2343-2357.

[5] Youssef, A. M., Pradhan, B., & Hassan, A. M. (2011). Flash flood risk estimation along the St. Katherine road, southern Sinai, Egypt using GIS based morphometry and satellite imagery. *Environmental Earth Sciences*, *62*(3), 611-623.

[6] Ruslan, F. A., Samad, A. M., Tajjudin, M., & Adnan, R. (2015, November). Flood water level prediction modeling using NNARX structure for Sg Pahang basin. In *2015 IEEE International Conference on Control System, Computing and Engineering (ICCSCE)* (pp. 548-551). IEEE.

[7] Olumide, B. A., Saidu, M., & Oluwasesan, A. (2013). Evaluation of best fit probability distribution models for the prediction of rainfall and runoff volume (Case Study Tagwai Dam, Minna-Nigeria). *International Journal of Engineering and Technology*, *3*(2).

[8] McCallum, I., Liu, W., See, L., Mechler, R., Keating, A., Hochrainer-Stigler, S., & Szoenyi, M. (2016). Technologies to Support Community Flood Disaster Risk Reduction. *International Journal of Disaster Risk Science*, *7*(2), 198-204.

[9] Yu, C. H., Luo, D., Ding, W., Cohen, J., Small, D., & Islam, S. (2015, October). Spatiotemporal asynchronous co-occurrence pattern for big climate data towards long-lead flood prediction. In Big Data (Big Data), 2015 IEEE International Conference on (pp. 865-870). IEEE.

[10] Liu, Y. B., & De Smedt, F. (2004). WetSpa extension, a GIS-based hydrologic model for flood prediction and watershed management. Vrije Universiteit Brussel, Belgium, 1, e108.

[11] Tehrany, M. S., Lee, M. J., Pradhan, B., Jebur, M. N., & Lee, S. (2014). Flood susceptibility mapping using integrated bivariate and multivariate statistical models. Environmental earth sciences, 72(10), 4001-4015.

[12] Botzen, W. J., Aerts, J. C., & van den Bergh, J. C. (2009). Dependence of flood risk perceptions on socioeconomic and objective risk factors. Water Resources Research, 45(10).