

# A Belief Rule Based Flood Risk Assessment Expert System using Real Time Sensor Data Streaming

Ahmed Afif Monrat  
Pervasive and Mobile Computing  
Laboratory  
Luleå University of Technology  
SE-931 87, Skellefteå, Sweden  
ahmmon-7@student.ltu.se

Raihan Ul Islam  
Pervasive and Mobile Computing  
Laboratory  
Luleå University of Technology  
SE-931 87, Skellefteå, Sweden  
raihan.ul.islam@ltu.se

Mohammad Shahadat Hossain  
Department of Computer Science and  
Engineering  
University of Chittagong  
Chittagong, Bangladesh  
hossain\_ms@cu.ac.bd

Karl Andersson  
Pervasive and Mobile Computing  
Laboratory  
Luleå University of Technology  
SE-931 87, Skellefteå, Sweden  
karl.andersson@ltu.se

**Abstract**—Among the various natural calamities, flood is considered one of the most catastrophic natural hazards, which has a significant impact on the socio-economic lifeline of a country. The Assessment of flood risks facilitates taking appropriate measures to reduce the consequences of flooding. The flood risk assessment requires Big data which are coming from different sources, such as sensors, social media, and organizations. However, these data sources contain various types of uncertainties because of the presence of incomplete and inaccurate information. This paper presents a Belief rule-based expert system (BRBES) which is developed in Big data platform to assess flood risk in real time. The system processes extremely large dataset by integrating BRBES with Apache Spark while a web-based interface has developed allowing the visualization of flood risk in real time. Since the integrated BRBES employs knowledge driven learning mechanism, it has been compared with other data-driven learning mechanisms to determine the reliability in assessing flood risk. The integrated BRBES produces reliable results in comparison to other data-driven approaches. Data for the expert system has been collected by considering different case study areas of Bangladesh to validate the system.

**Keywords**—Belief Rule Base, Flood risk assessment, Uncertainty, Expert systems, Sensor data streaming, Big data.

## I. INTRODUCTION

Flood is one of the pervasive hazards with catastrophic consequences. It poses a greater threat to all aspects of sustainability including social, economic and environmental in comparison with other natural disasters such as earthquakes, landslides, cyclone and volcanic eruptions [1]. It is not only responsible for the loss of human lives but also inundates residential and commercial properties harming local economies, causes disruption in national transport infrastructure, agriculture, electrical grid. Furthermore, severe water contamination and health hazards can occur due to flooding, for example, a huge tsunami hit Japan in 2011 and a part of the coastline was flooded by the sea water which caused massive leakage in nuclear plants that ended up contaminating waters with chemicals and hazardous substances [2]. It is estimated that the yearly cost to the world

economy because of flooding is around 50-60 billion US dollars [3]. As per an examination by the United Nations (UN), approximately 22,800 lives are affected significantly by flood in a year and caused an estimated damage of US\$ 136 billion to the Asian economy [4]. It is not possible to avoid flood risk entirely. Several factors such as topographical, meteorological, geological, river characteristics and human activities are responsible for flooding [5]. These factors can be categorized as quantitative and qualitative in nature. For instance, financial loss can be measured in a quantitative way while health condition should be expressed in a qualitative way. There are different types of uncertainties associated with each factor such as ignorance, incompleteness, ambiguity, imprecision and vagueness [6] as shown in Table I.

TABLE I. FACTORS OF FLOOD RISK ASSESSMENT AND THEIR RELATED UNCERTAINTIES

Factors	Types of Uncertainty	Discussion
Water Level	Incompleteness	Refers to the flood water level.
Accommodation Problem	Vagueness	Problem with accommodation in flood affected area.
Availability of Transport	Incompleteness, Imprecision	Refers to the availability of transportation in flood affected area.
Transportation Delay	Ignorance	Delay in transportation due to flood water.
Road Damage	Imprecision	Percentage of road damage due to flood.
Financial Condition	Vagueness	Financial condition of the flood affected people.
Social Condition	Vagueness, Ignorance	Refers to the social condition of flood affected area.
Damage of Crops	Incompleteness	damage of crops in the flood affected area.
Loss of Cattle	Imprecision, Incompleteness	Percentage of cattle were died during the flood.
Casualties	Imprecision	How many people died during flood.
Health Hazards	Ignorance, vagueness	Refers to health issues due to water borne out- breaks
Water Contamination	Incompleteness, Imprecision	Refers to the scarcity of fresh water

These uncertainties can be caused by anomalous sensor data or lack of human knowledge [7]. For example, the erroneous and misleading nature of sensor data makes the

prediction highly unreliable. Battery power, computational and memory capacities, as well as communication bandwidths are the resource constraints which are responsible for missing, redundant or inconsistent data [8]. Besides, sensors have vulnerability issues regarding the malicious attack, for example, eavesdropping, indiscretion, integrity violation, denial of service and black hole attack [9]. Similarly, human ignorance to define qualitative data can lead towards uncertainties due to vagueness and ignorance, for example, “social condition” can be expressed in terms of high, medium and low and the answer varied in a wide range as well. Therefore, it is mandatory to develop a reliable system that can address all sorts of uncertainties to assess flood risks.

In this paper, a belief rule-based expert system (BRBES) has been developed which has the capability of processing heterogeneous data with various types of uncertainties. It works like a single integrated framework to assess the consequences or risks of the flood in an area. In addition, to get better prediction and accuracy, BRBES has been integrated with a Big data Analytics platform because the traditional approach is not capable of handling voluminous dataset as well as performing complex mathematical computations.

The remainder of this article is structured as follows: Section two covers background and literature review, an overview of methodology and system framework is given in Section 3, while Section 4 describes the implementation along with the architecture of proposed BRBES; Section 5 presents the results and discussion, while Section 6 concludes the paper and indicates future work.

## II. RELATED WORK

There are different existing methods which are used by various research groups and organizations throughout the world for accessing flood risk. Kussul et al. [10] described a general approach for flood monitoring and risk assessment based on heterogeneous geospatial data acquired from multiple sources to provide disaster hazard mapping. The method derives a probability density function of the disaster which is estimated based on the analysis of heterogeneous geospatial data and it follows a sensor web approach which provides standardized services for acquiring and planning observations, establishing alerts and warning, etc. The proposed approach is being implemented within the Namibia SensorWeb Pilot Project.

The National Flood Insurance Program (NFIP) uses a hydrological method developed in the 1960's by U.S Department of Housing and Urban Development (HUD) of USA to assess flood risk for insurance purposes [11]. This method derives a water surface elevation probabilistic function to evaluate the flood risk in a geographical area. It transforms the probabilistic function as a function of depth inundation using a model of damage. This probability function computes the average annual loss. NFIP also has a more comprehensive model for risk analysis that can evaluate site-specific probabilities to represent the risk. It considers the

performance and reliability of flood protection including the effect of their failure on flooding.

For Addressing Uncertainty, a web-based belief rule-based expert system for assessing flood risk was introduced by Rafiul et al. [12], where he proposed an expert system by combining belief rule base with the capability of reading data and generating web-based output. Moreover, the author developed a generic RESTful API which makes the system more robust and highly accessible by providing a layer of abstraction for the users who want to use the expert system without knowing much about the underlying algorithm. However, this approach is not scalable because the system is not capable to work with sensor data streaming as it provides a huge amount of data.

Nowadays, the crisis response teams of a country have turned their interest to explore the potentialities of Big Data in managing disaster risks such as flooding [13]. The reason for this is that during flooding, crisis response teams need to make decisions based on the huge amount of incomplete and inaccurate information. A group of scientists (De Groeve, Kugler and Brakenridge) built a real-time map of the location, timing, and impact of floods by combining information related to flood from Twitter and satellite observations [14]. In this paper, the authors proposed a space-based river monitoring framework that can provide a systematic and impartial way to predict floods. The results obtained with passive microwave remote sensing considering flood extent and flood start and duration. Optical high-resolution satellite imagery has been used for validating extent measurements. It allows to have a quantitative assessment of the size of the floods, either relative to previous years or in absolute terms. The technique has been demonstrated in Southern Africa during the floods of 2009.

An early warning system has been developed by the New South Wales state emergency service in Australia [15]. This research introduced an interactive web-GIS tool as a prototype which is based on open-source geospatial software and technologies to perform risk analysis of natural hazards, especially for floods and landslides. It takes meteorological dataset such as data from the floodplain, historical data information from various databases to perform predictive analysis of floods in the different region. The users can monitor and analyze the areas at risk by importing necessary maps within this platform.

A social media analytics platform named Floodtags was deployed for extracting information from Twitter. It has the functionality to perform filtering, visualization and mapping social media content based on location and keywords [16]. Besides, it also provides a service through microwave satellite observations for identifying inundated areas rapidly. The approach has been used in Philippines and Pakistan as case studies which later proved to be a great success monitoring large floods in densely populated areas.

All these flood risk assessment methods described above address different issues of flood risks for specific scenarios. While some of these methods explicitly consider uncertainty features such as web-based BRB expert system, some frameworks are well equipped to provide scalable solutions

while working with huge amount of complex data. Therefore, it can be argued that none of these solutions can be considered as the best approach while dealing with both uncertainties as well as voluminous sensor streaming data. For this reason, this research emphasizes on integrating big data platform with BRBES to mitigate the constraints mentioned above and introduce a new learning model which will provide scalability and availability while addressing all sorts of uncertainties.

### III. METHODOLOGY AND INFERENCE PROCEDURES

This section discusses the procedure of knowledgebase construction and inference mechanism of the Belief Rule Based Expert System (BRBES). In addition, the framework, challenge and opportunities of Big data platform in flood risk assessment is also presented.

#### A. Knowledge Base Construction of BRBES

A Belief Rule Base is a knowledge representation schema. It is an extension of IF-THEN rules and capable of representing uncertainties [17]. There are two main parts of a Belief Rule Based Expert System (BRBES): a knowledge base and inference engine as shown in Fig. 3. The main functionality of Belief Rule Base is to build initial rule base whereas evidential reasoning is used as an inference engine. Different learning parameters such as rule weight, antecedent attribute weight and belief degrees are associated with a belief rule [18][19][20][21]. A belief rule is represented by Equation (1).

$$\left\{ \begin{array}{l} \text{IF} \\ (X_1 \text{ is } X_1^k) \wedge (X_2 \text{ is } X_2^k) \wedge \dots \wedge (X_{T_k} \text{ is } X_{T_k}^k) \\ \text{THEN} \\ \{(P_1, \bar{\beta}_{k1}), (P_2, \bar{\beta}_{k2}), \dots, (P_N, \bar{\beta}_{kN})\}, \left( \left( \sum_{n=1}^N \bar{\beta}_{kn} \leq 1 \right) \right), \\ \text{with rule weight } 0 \leq \theta_k \leq 1, \\ \text{and attribute weight } \delta_1^k, \delta_2^k, \dots, \delta_T^k \geq 0 \\ \text{satisfying } \sum_{i=1}^{T_k} \delta_i^k = 1 \end{array} \right. \quad (1)$$

where  $X_1, X_2, \dots, X_{T_k}, T_k \in \{1, 2, \dots, T\}$  represents the antecedent attributes used in the  $k$ th rule and  $P_1, P_2, \dots, P_N$  are the referential values of the consequent attribute  $X_i$  where  $\bar{\beta}_{ki}$  is the belief degree.

For example, if onset rainfall is “High” and prolonged rainfall is “Low” then Meteorological factor is:  $\{(High, 0.8), (Medium, 0.2), (Low, 0)\}$

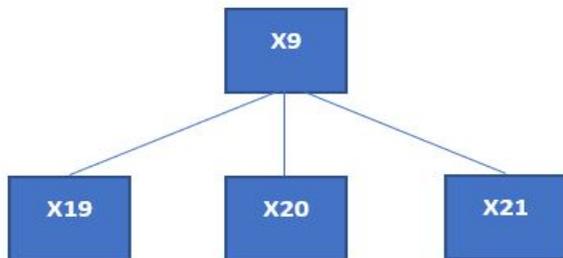


Fig. 1. A SIMPLE BRB TREE DIAGRAM

From Fig. 1 we can observe that, X19, X20 and X21 are the antecedent attributes, while “X9” is the consequent attribute. The referential values associated with the antecedent attributes comprise high, medium and low. The illustrated rule is complete since the total sum of the degree of beliefs stands at '1'. “X9” BRB consists of three antecedent attributes, each with three referential values, namely high, medium and low. As a result, the “X9” BRB consists of 27 rules according to (2) [21].

$$L = \prod_{i=1}^T J_i \quad (2)$$

where

$L$  = The number of rules in a BRB

$J_i$  = Referential values of the  $i^{\text{th}}$  antecedent attribute

#### B. Inference Mechanisms of BRBES

Evidential Reasoning (ER) is a decision analysis procedure which can handle heterogeneous data [22]. ER can handle different types of uncertainties such as incompleteness, ignorance, imprecision, vagueness exists in data [23]. The inference mechanisms using Evidential Reasoning (RIMER) approach consists of four steps including input transformation, rule activation calculation, belief update, and rule aggregation as shown in Fig. 2 [24]. During input transformation, the input data gets distributed over the referential values of the attribute of a rule, which is called matching degree. Then the rules are called packet antecedent and they become active. This matching degree is used to calculate the activation weight of each rule. While calculating the referential values of the consequent attribute, it is interesting to observe that each rule has different weight. When an input data for any of the antecedent attribute is ignored or missing, the belief degree associated with each rule in the rule base should be updated. This is done by the procedures mentioned in [25]. The output of the BRB system is fuzzy, meaning that qualitative values can be transformed into crisp values, i.e. quantitative data, by assigning utility scores to each referential value of the consequent attributes.

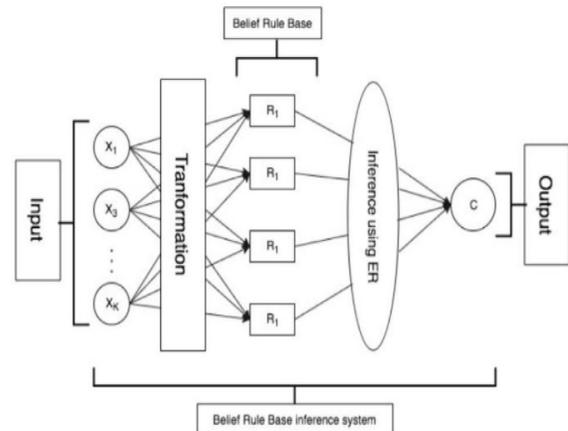


Fig. 2. SEQUENCE OF BRBES INFERENCE MECHANISMS

### C. Big Data Analytics Platform

This segment will explain the scope and framework of Big Data that can bring new opportunities to improve the way flood risk assessment is planned and executed. By definition, when a collection of dataset becomes large and complex which exceeds the processing capability of conventional data management techniques and system, it's called Big data [26]. This growing phenomenon is capable of collecting, storing and processing huge data stream coming from three major sources, including people, machines and organizations. For example, data coming from various sensors, cellphones, stock exchange, geospatial data, satellite data, security footage and YouTube videos, organizational and social media data etc. can be processed via this robust framework. Big data provides features such as scalability, reliability, consistency, high availability, replication and security which eliminates the possibility of single server failure, data corruption, scarcity of storage etc. [27]. In this research, we studied the possibilities of Big data platform regarding the context of flood risk assessment. Big Data technologies can be used to monitor and determine the people, exposed to the risks of flooding in real time [28]. This could be achieved by analyzing and processing sensor data streams coming from various sources as well as data collected from other sources such as Twitter, Facebook, satellite geospatial data and also from disaster organizations [29]. With Big Data, records of previous flood incidents such as fatality, the number of damaged properties, rainfall during that period, infrastructures of the areas including coastal areas as well as cities and drainage system can be analyzed properly [30]. It can also pick out the specific mobility support or resources that are needed by the inhabitants of a flood affected area. Hence, identifying population hotspot gets easier with Big Data to provide real-time alarm and warnings to the residents when a disaster approaching.

In order to facilitate better flood risk assessment, we have used a hybrid Spark model with Hadoop ecosystem. Hadoop scales huge dataset using commodity systems [31]. Spark provides in-memory processing for real time data. When Apache Spark's ability, i.e. high processing speed, advanced analytics and multiple integration support combines with Hadoop's low-cost operation on commodity hardware, it gives the best results. In this approach, Spark will run on top of Hadoop Distributed File System (HDFS) to leverage it's potential for distributed replicated storage. Furthermore, Spark applications for real time processing can be scheduled by YARN. This architecture will provide us excellent storage capability along with complex computation possibilities through Spark machine learning libraries. Fig. 3 shows us how Spark can utilize Hadoop through HDFS for reading and storing data while YARN for resource allocation and MapReduce for optional processing.

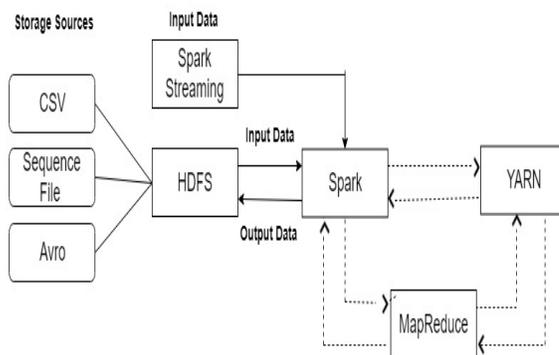


Fig. 3. ARCHITECTURE OF HYBRID SPARK MODEL

### IV. BRBES TO ASSESS FLOOD RISK

This section describes the system architecture and implementation techniques of the proposed BRBES system integrated with Big data analytics platform for assessing flood risks.

#### A. System Architecture

We can consider the system architecture in Fig. 4 to assess flood risk using Big Data. The system has three major parts: data source, a hybrid SPARK distributed server and a web interface. The data source will provide data to the server, which will perform computation regarding flood risks by using Big Data analytics platform, such as Apache Spark, where BRBES will train and validate the model using knowledge base approach and finally through a web service it will visualize the risk in a web-based user interface.

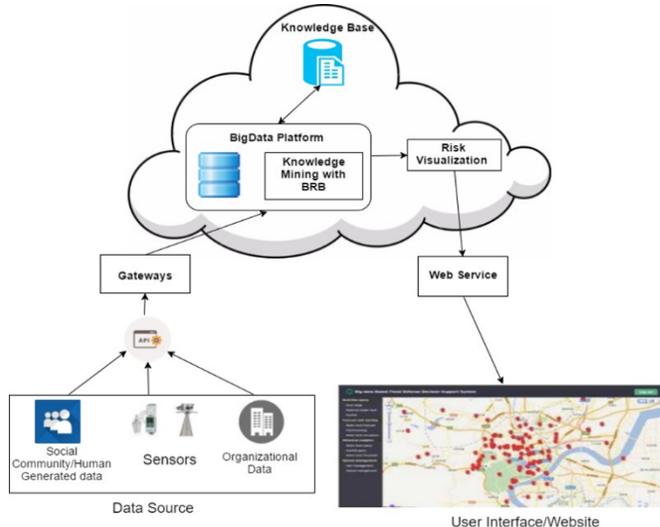


Fig. 4. SYSTEM ARCHITECTURE OF INTEGRATED BRBES WITH BIG DATA FRAMEWORK

It has three major data sources including human-generated data (Twitter, web traffic), sensors and data that are coming from different organizations. Real time data was provided by Flood Forecasting and Warning Centre, Bangladesh Water Development Board (BWDB) and for social media response for flooding in any specific region, twitter data streams have been considered. All these data will be forwarded towards a server through an API gateway. The Big Data platform will store this huge amount of data in

different nodes by replicating several times to ensure the integrity of data providing fault tolerance, high availability and scalability. YARN will be used as resource management layer that will distribute the computing tasks into different nodes. Then these data will be analyzed using BRB inference engine to get rid of the uncertainties associated with the data as well as to produce the risk assessment. The users can view the final assessment in a website for the specific region. In addition, clients can also monitor the factors like water level, rainfall and moisture of soil. The system will produce real time assessment by computing stream of data coming from different sources. Therefore, crisis management teams, authorities and people will be able to take proper precautions before the flood occurs.

### B. Implementation of Integrated BRBES Framework

Initially, data are coming from different sources such as sensors or twitter stream. HDFS then stores all these information in different nodes and performs replication in order to maintain availability. For scalability issues, we have used Amazon EC2 clusters which is capable of handling data traffic through load balancing on demand. Spark Engine initiates the scheduler (YARN) to distribute the jobs among the task nodes.

This integrated BRBES framework developed here adopt a four-layer architecture model, including the Data Management Layer, the Application layer, the Application Programming Interface (API) layer, and the Interface layer as shown in Fig. 5.

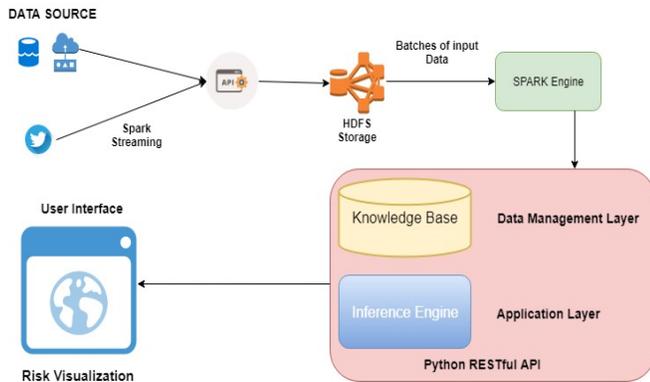


Fig. 5. ARCHITECTURE DIAGRAM OF INTEGRATED BRBES FRAMEWORK

The Data Management layer creates initial rule base using the data collected from different sources through API gateway.

The initial BRB is generated in the data management layer which is the knowledge base of this integrated system. Fig. 6 shows a BRB framework that is developed to construct the knowledge base of the system.

The Application layer consists of data access procedures as well as inference procedures of BRB as discussed in the previous section. This phase will rigorously train and validate the data to extract valuable insights, in this case it is the risk of flood in any particular region or river station. Data ingestion and processing is the main focus of this inference phase. HDFS contains all the data in CSV format including the Spark streaming real time twitter feedbacks. We will get only those data relevant to flood filtering those data with twitter codified hash tagging such as #FloodRelief, #FloodAlert etc.

Spark will use its Resilient Distributed Datasets (RDD), which is used to hold data objects in memory and transformations that are performed on the datasets [32]. With read-only-mode and fault tolerant features it can allow users making intermediate results to persist in the memory. Moreover, it is possible to control data partition for optimizing data placement as well as manipulating them using a rich set of operators. Spark implements a machine learning library known as MLlib that supports all the common features of machine learning problem settings including binary classification, regressing, clustering, optimization and collaborative filtering. In this phase, Spark will run “spark-submit” command which will take initial BRB as input from data management layer and run the inference mechanism on it.

A dynamic tree traversal algorithm has been deployed because of the multi-level architecture of initial BRB [12]. The beauty of the above described tree traversal algorithm is that it is generic and can be applicable for any BRB tree from any domain. No matter for which domain the BRB tree is created for, this algorithm can traverse the whole tree from bottom to top and can generate the result of the top node by calculating the subtrees step by step as shown in Fig. 6.

The Big data features enables BRBES with faster processing of huge amount data (Tera/Peta Bytes) and extracts the true value, that is the risks associated with it. In traditional BRBES system it would have taken days or even months to process large unstructured datasets while this architecture can analyze these data within hours. Besides, this framework can produce efficient outcome regardless of batch processing or in memory processing as it is a hybrid model of both Hadoop and Spark ecosystem.

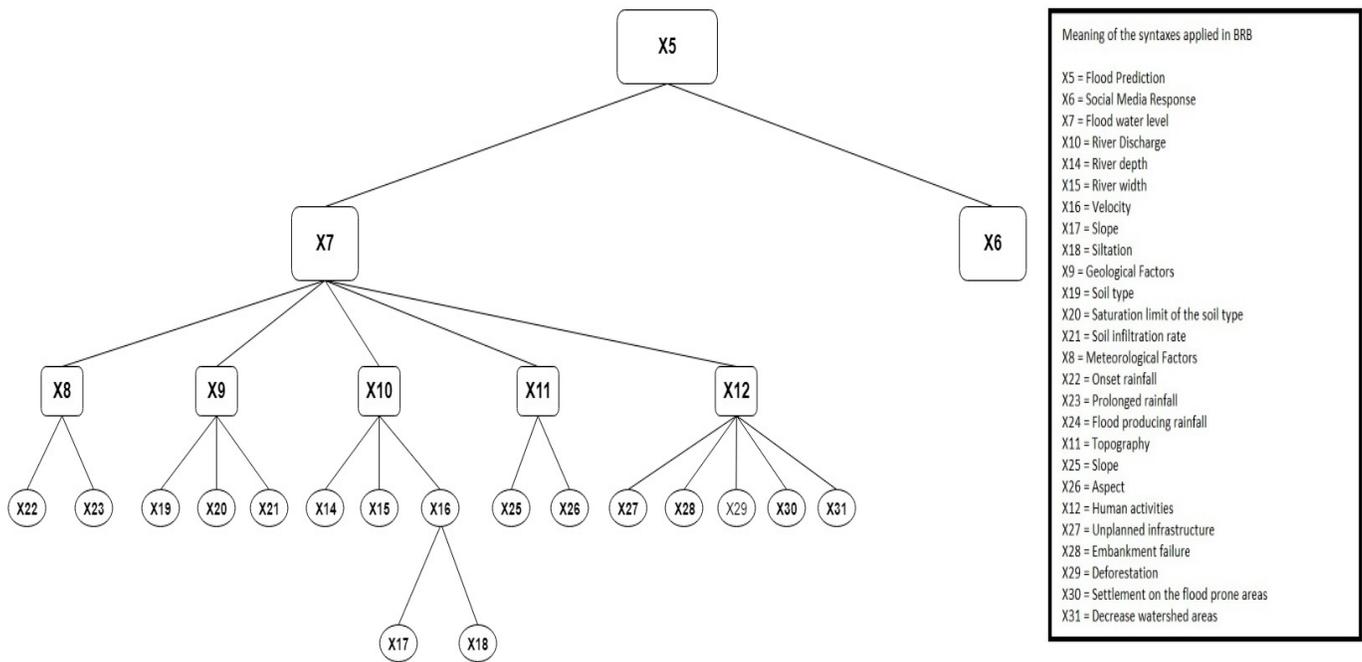


Fig. 6. BRB FRAMEWORK TO ACCESS THE FLOOD RISK

The predicted risk assessment along with the input factors will be stored in a distributed database management system (MongoDB). In this research a web-based tool, named FloodMap has been introduced to visualize the risks in a web interface. It will help the users to monitor the flood risks in specific areas. The integrated Spark-BRBES tool has been developed using python and the user interface was constructed using REST complaint webservices as they are more lightweight than SOAP and easier to implement by PHP and JavaScript. Fig. 7 shows the Graphical User Inter-face (GUI) of the proposed FloodMap system which has been developed by using Hypertext Markup Language (HTML), Cascading Style Sheets (CSS) and PHP as server-side script.

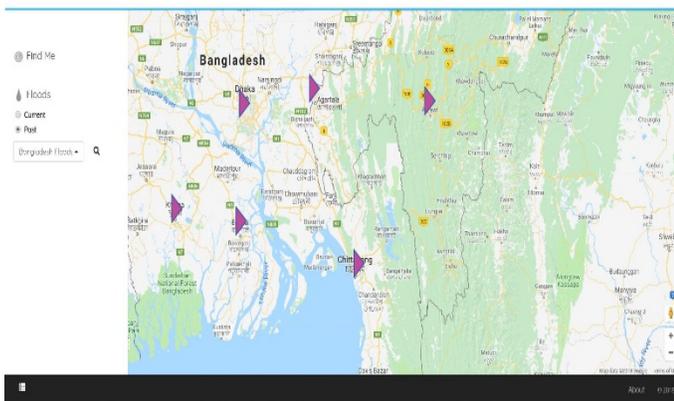


Fig. 7. FLOOD RISK VISUALIZATION USING FLOODMAP TOOL

## V. RESULTS AND DISCUSSION

In order to check the accuracy and the reliability of the system's results to assess flood risk, data samples were collected from the Flood Forecasting and Warning Centre, Bangladesh Water Development Board (BWDB) and for social media response for flooding, Twitter data streams have been considered. SQL data generator tool has been used to generate realistic test data on volumes based on the sample dataset. It is a data replication tool used for User Acceptance Testing (UAT) Environments [33]. Approximately, 10 GB data was produced using this tool.

This research studied different performance metrics such as Sensitivity, Root Mean Square Error (RMSE) and Area Under Curve (AUC) to contrast the reliability among different machine learning approaches integrated with Big data platform. A comprehensive study with performance evaluation has been performed to demonstrate how knowledge driven approach (BRBES) is performing compare to other data driven model, for example ANN, linear regression, decision tree and random forest. The knowledge driven tradition emphasizes rules and logics as a tool for representing beliefs while the data driven approach only focused on observed data and does not use logic as a modelling tool [34]. However, the emergence of fuzzy reasoning and BRBES has blurred the gap between these two traditions by putting rules for knowledge representation while non-linear functions represents data [35]. As fuzzy logic is not capable of handling all kinds of uncertainty, BRBES is proven to be the most efficient approach while dealing with uncertainty regarding risk assessments [36].

TABLE II. FLOOD RISK ASSESSMENT BY BRBES, ANN, LINEAR REGRESSION, RANDOM FOREST, DECISION TREE AND EXPERT

SL. NO (1)	X8 (2)	X9 (3)	X10 (4)	X11 (5)	X12 (6)	X7 (7)	X6 (8)	BRBES (9)	ANN (10)	Linear Regression (11)	Random Forest (12)	Decision Tree (13)
1	0.808	0.512	103	0.368	46	349	34	85.25	81.23	70.41	64.13	62.16
2	0.661	0.443	80	0.261	64	372	47	91.41	87.57	74.42	46.23	40.32
3	0.566	0.450	86	0.364	30	315	95	22.33	34.11	23.12	34.78	30.12
4	0.894	0.551	83	0.476	41	364	15	39.34	44.89	32.22	23.67	27.39
5	0.543	0.595	100	0.293	45	312	95	63.28	45.27	48.12	84.05	77.34

TABLE III. COMPARISON OF RELIABILITY AMONG FIVE SYSTEMS

Model (1)	Sensitivity % (2)	Specificity % (3)	RMSE (4)	AUC (5)	95% CI (6)
Decision Tree	80.34	66.96	0.3602	0.752	0.634 – 0.870
Random Forest	83.33	78.89	0.3273	0.800	0.648 – 0.915
Linear Regression	89.13	85.37	0.3023	0.826	0.725 – 0.927
ANN	87.11	89.23	0.2657	0.874	0.789 – 0.959
BRBES	91.31	92.68	0.2347	0.895	0.816 – 0.973

The results of the BRBES have been compared with different machine learning approaches such as Artificial Neural Networks (ANN), Linear Regression, Decision Tree and Random Forest. The machine learning methods were tested in Apache Spark 2.2.0 environment. Table II presents different factors for flood risk assessment (column 2 through 8) as well as BRBES and the mentioned machine learning tool’s produced results (column 9 through 13). To compare and evaluate the performance of a predictive model like BRBES, Receiver Operating Characteristics Curves (ROC) are widely employed. In addition, different performance matrices including Sensitivity, Specificity, Area Under Curve (AUC), Confidence Interval and Root Mean Square Error (RMSE) are also frequently employed.

that of other methods. Further, it can also be observed that the range of confidence interval of BRBES as shown in column 6 is the highest among all the methods. Column 4 of Table III also illustrates that the RMSE value of BRBES is the lowest among other methods. This means that BRBES's results produce less error than that of other methods. Therefore, a correlation among the matrices such as AUC, CI and RMSE can easily be observed; and hence, BRBES is performing better than that of other methods. BRBES is also enjoying the highest sensitivity and specificity values in comparison to other methods as can be seen from Column 2 and 3 of Table III. This means that BRBES can correctly identify the risk for flood prone areas.

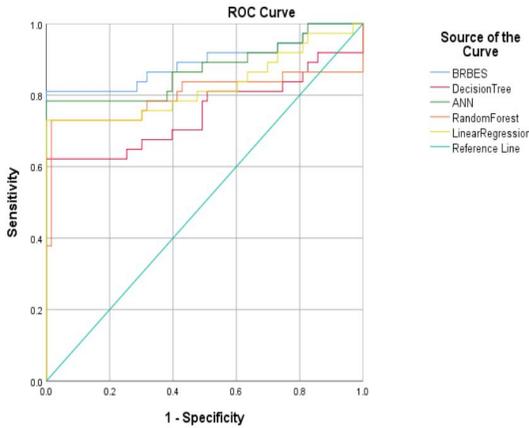


Fig. 8. COMPARISON OF RESULTS OF BRBES, LINEAR REGRESSION, RANDOM FOREST, DECISION TREE AND ANN USING ROC CURVES

The results of the mentioned matrices associated with the BRBES and various machine learning tools are illustrated in Table III. Fig. 8 illustrates the ROC of the five methods. ROC graphs are useful and widely popular for organizing and selecting classifiers by visualizing their performance [36]. AUC is associated with the ROC and the greater its value indicates the more accuracy and the reliability of the predictive models. From an investigation of the AUC values of column 5 of Table III, it can be observed that BRBES is enjoying the highest value and hence, its generated results are better than

VI. CONCLUSION AND FUTURE WORK

This research explores the challenges, opportunities and methods, required to leverage the potentiality of Big Data and BRBES to assess the risk of flooding using sensor data streaming. The integrated BRBES framework proves to become more efficient comparing to other machine learning approaches. The proposed model can address different types of uncertainties as well as it can perform inference mechanisms on voluminous dataset to extract the precise assessment for flood risk. The RESTful API also makes the system more robust and accessible. The FloodMap tool is capable of visualizing the flood risk in real time. The government as well as the authority, crisis response team and general people can take immediate precautions to mitigate the sufferings to a great extent. Although the system is providing good assessment, the dataset regarding the computation is not quite large. It will be interesting to see how it performs while processing petabytes or yottabytes of data. Security is also an important aspect as large scale of data coming from different IoT (Internet of Things) device..

ACKNOWLEDGMENT

This research has been supported by Pervasive Computing and Communications for Sustainable Development (PERCCOM) and the Swedish Research Council under grant 2014-4251. PERCCOM is a joint master’s degree program funded by a grant from the European Union Erasmus Mundus program [37]. The authors would like to acknowledge the European Union.

## REFERENCES

- [1] L.J. Hill, R.S.J. Sparks, and J.C. Rougier, "Risk assessment and uncertainty in natural hazards." Risk and uncertainty assessment for natural hazards, edited by: Rougier, JC, Sparks, RS J., and Hill, LJ, pp. 1–18.
- [2] T. Tominaga, M. Hachiya, H. Tatsuzaki, M. Akashi, "The accident at the Fukushima Daiichi nuclear power plant in 2011," *Health physics*, 106(6), pp. 630–637, 2014.
- [3] A.B. Shrestha, S.H. Shah, and R. Karim, "Resource manual on flash flood risk management," *Internat. Centre for Integrated Mountain Development, ICIMOD*, 2008.
- [4] M.S. Shrestha and K. Takara, "Impacts of floods in south Asia," *Journal of South Asia Disaster Study*, 1(1), pp. 85–106, 2008.
- [5] N.N. Kourgialas and G.P. Karatzas, "Flood management and a GIS modelling method to assess flood-hazard areas—a case study," *Hydrological Sciences Journal—Journal des Sciences Hydrologiques*, 56(2), pp. 212–225, 2011.
- [6] R. Ul Islam, M.S. Hossain, and K. Andersson, "A novel anomaly detection algorithm for sensor data under uncertainty," *Soft Computing*, 22(5), pp. 1623–1639, 2018.
- [7] P. Ramuhalli, G. Lin, S.L. Crawford, B.A. Konomi, J.B. Coble, B. Shumaker, and H. Hashemian, "Uncertainty quantification techniques for sensor calibration monitoring in nuclear power plants," (No. PNNL-22847 Rev. 1). Pacific Northwest National Laboratory (PNNL), Richland, WA (US), 2014.
- [8] M.S. Hossain, S. Rahman, A.L. Kor, K. Andersson, and C. Pattinson, "A Belief Rule Based Expert System for Datacenter PUE Prediction under Uncertainty", *IEEE Transactions on Sustainable Computing* 2(2):140–153, 2017.
- [9] R. Ul Islam, K. Andersson, and M.S. Hossain, "A Web Based Belief Rule Based Expert System to Predict Flood," 17th International Conference on Information Integration and Web-based Applications & Services (iiWAS 2015), 2015.
- [10] N. Kussul, D. Mandl, K. Moe, J.P. Mund, J. Post, A. Shelestov, S. Skakun, J. Szarzynski, G. Van Langenhove, and M. Handy, 2012. Interoperable infrastructure for flood monitoring: SensorWeb, grid and cloud," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(6), pp. 1740–1745, 2012.
- [11] S.A. Mohammadi, M. Nazariha, and N. Mehrdadi, "Flood damage estimate (quantity), using HEC-FDA model. Case study: the Neka river," *Procedia Engineering*, 70, pp. 1173–1182, 2014.
- [12] M.R.S. Hridoy, R. Ul Islam, M.S. Hossain, and K. Andersson, "A Web Based Belief Rule Based Expert System for Assessing Flood Risk," The 19th International Conference on Information Integration and Web-based Applications & Services (iiWAS2017), 2017.
- [13] T. Papadopoulos, A. Gunasekaran, R. Dubey, N. Altay, S.J. Childe, and S. Fosso-Wamba, "The role of Big Data in explaining disaster resilience in supply chains for sustainability," *Journal of Cleaner Production*, 142, pp. 1108–1118, 2017.
- [14] Z. Kugler and T. De Groeve, "The global flood detection system," Office for Official Publications of the European Communities, Luxembourg, 2007.
- [15] A. Zenger, "Examining GIS decision utility for natural hazard risk modelling," *Environmental Modelling & Software*, 17(3), pp. 287–294, 2002.
- [16] B. Jongman, J. Wagemaker, B.R. Romero, and E.C. de Perez, "Early flood detection for rapid humanitarian response: harnessing near real-time satellite and Twitter signals," *ISPRS International Journal of Geo-Information*, 4(4), pp. 2246–2266, 2015.
- [17] M.S. Hossain, S. Rahaman, R. Mustafa, and K. Andersson, "A belief rule-based expert system to assess suspicion of acute coronary syndrome (ACS) under uncertainty," *Soft Computing*, pp. 1–16. In Press, 2017.
- [18] T. Mahmud and M.S. Hossain, "An evidential reasoning-based decision support system to support house hunting," *International Journal of Computer Applications*, 57(21), pp. 51–58, 2012.
- [19] M.S. Hossain, I.B. Habib, and K. Andersson, "A belief rule based expert system to diagnose dengue fever under uncertainty," *Computing Conference*, 2017, pp. 179–186, 2017.
- [20] M.S. Hossain, K. Andersson, and S. Naznin, "A belief rule based expert system to diagnose measles under uncertainty," *World Congress in Computer Science, Computer Engineering, and Applied Computing (WORLDCOMP'15): The 2015 International Conference on Health Informatics and Medical Systems*, 2015.
- [21] M.S. Hossain, M.S. Khalid, S. Akter, and S. Dey, "A belief rule-based expert system to diagnose influenza," 2014 9th International Forum on Strategic Technology (IFOST), 2014.
- [22] K. Andersson and M.S. Hossain, "Smart risk assessment systems using belief-rule-based DSS and WSN technologies," 2014 4th International Conference on Wireless Communications, Vehicular Technology, Information Theory and Aerospace & Electronic Systems (VITAE), 2014.
- [23] J.B. Yang, J. Liu, J. Wang, H.S. Sii, and H.W. Wang, "Belief rule-base inference methodology using the evidential reasoning approach-RIMER," *IEEE Transactions on systems, Man, and Cybernetics-part A: Systems and Humans*, 36(2), pp. 266–285, 2006.
- [24] M.S. Hossain, F. Ahmed, F. Tuj-Johora, and K. Andersson, "A Belief Rule Based Expert System to Assess Tuberculosis under Uncertainty," *Journal of medical systems* 41:43, 2017.
- [25] T. Mahmud and M.S. Hossain, "An Evidential Reasoning-based Decision Support System to Support House Hunting", *International Journal of Computer Applications*, Vol. 57, No. 21, pp. 51–58, 2012.
- [26] C.P. Chen and C.Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data," *Information Sciences*, 275, pp. 314–347, 2014.
- [27] J. Baker, C. Bond, J.C. Corbett, J.J. Furman, A. Khorlin, J. Larson, J.M. Leon, Y. Li, A. Lloyd, and V. Yushprakh, "Megastore: Providing scalable, highly available storage for interactive services," *CIDR*, Vol. 11, pp. 223–234, 2011.
- [28] A.A. Monrat, R. Ul Islam, M.S. Hossain, and K. Andersson, "Challenges and Opportunities of Using Big Data for Assessing Flood Risks," *Applications of Big Data Analytics*, pp. 31–42, Springer, Cham, 2018.
- [29] M.S. Hossain, and R. Mustafa, "Resolving geo-spatial semantic conflicts—an interoperability issue," *Humanomics*, 23(2), pp. 102–109, 2007.
- [30] T.H. Dewan, "Societal impacts and vulnerability to floods in Bangladesh and Nepal," *Weather and Climate Extremes*, 7, pp.36–42, 2015.
- [31] K. Venkatram and M.A. Geetha, "Review on Big Data & Analytics—Concepts, Philosophy, Process and Applications", *Cybernetics and Information Technologies*, 17(2), pp. 3–27, 2017.
- [32] M. Zaharia, M. Chowdhury, T. Das, A. Dave, J. Ma, M. McCauley, M.J. Franklin, S. Shenker, and I. Stoica, "Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing," *Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation*. USENIX Association, 2012.
- [33] R Ltd. SQL Data Generator - Data Generator For MS SQL Server Databases. [online] Red-gate.com. Available at: <https://www.red-gate.com/products/sql-development/sql-data-generator/index>. [2018-08-06].
- [34] D. Dubois, P. Hájek, and H. Prade, "Knowledge-driven versus data-driven logics. *Journal of logic*," *Language and information*, 9(1), pp. 65–89, 2000.
- [35] S.W. Khor, M.S. Khan, and K.W. Wong, "Fuzzy inferencing using single-antecedent fuzzy rules," *International Journal of Fuzzy Systems*, 8(2), pp. 65–75, 2006.
- [36] T. Fawcett, "ROC graphs: Notes and practical considerations for researchers," *Machine learning*, 31(1), pp. 1–38, 2004.
- [37] A. Klimova, E. Rondeau, K. Andersson, J. Porras, A. Rybin, and A. Zaslavsky, "An international Master's program in green ICT as a contribution to sustainable development," *Journal of Cleaner Production*, 135, pp. 223–239, 2016.