A Novel Hybrid Supervised Machine Learning Model for Real Time Risk Assessment of Floods using Concepts of Big Data

[1]TEGIL J JOHN,

[1]Research Scholar, Department of Computer Science, Kaamadhenu Arts and Science College, Sathyamangalam, Erode (D.t), Tamil Nadu-638 503.

[2]DR.R. NAGARAJ

[2]Associate Professor, Department of Computer Science, Kaamadhenu Arts and Science College,Sathyamangalam, Erode (D.t), Tamil Nadu-638 503.

DOI 10.5281/zenodo.6553253

Abstract: *Amongst various natural catastrophes, flood is one of the most destructive natural disasters, which has severe impact over the economy of a country and its people. Modelling and analyzing the flood is a complex task. The advancement in technologies helps to predict the impact of floods through an essential metric, namely, risk assessment, which drasticallyreducesthelossofhumanlife,propertydamage,throughpreventivemeasurestaken well in advance. Risk assessment modelling refers to combinatorial development of identification and assessment of potential for occurrence of an event which causes a negative impactonanentityofinterest.Withrecentadvancesindataacquisitionandarchivalmethods, concepts of big data have been a great boon to risk assessment development. It is primarily duetothefactthat,theaccuracyofriskassessment(RA)reliesonthevolumeofhistoricaldata analysed. Based on this, a risk assessment model is designed as a hybrid model using differentialevolutionandadaptiveneurofuzzyinferencesystemtoassessriskinrealtime.The performance ability of proposed hybrid model is compared with conventional ANFIS and neural network models by analysing the rainfall status in India. Based on the monthly, annual andseasonaldata,theriskassessmentisperformedthroughvariousfactorslikewetnessindex, land slope angle, stream power, stream density, rainfall, curvature and distance. Data from the expert systems are collected by analysing various case study areas from India to validate the performance of proposed hybridsystem.*

Keywords: Flood risk assessment, Prediction models, Big Data, Machine learning, Hybridization, Prediction accuracy

I. INTRODUCTION

Worldwidefloodisoneamongthesignificantnaturaldisasterswhichtakesawaynumerous lives, produces intense property damage, collapses the socio-economic factors, creeps up the ecosystem and paralyzes transportation. Every year on an average, 200 million people get affected,whichresultsintoa$95billioneconomiclossaroundtheworld[1].Duetothefloods, almost50%ofeconomiclossisclaimedbyAsiancountries,whichcompriseshumancasualties on an average of 90%. India has witnessed continuous floods recently, especially in North and SouthIndiawhichhasresultedinhugefinanciallossandhumanlives.Table.1depictstheflood statistics of India from 2013 to 2020.

Table 1. Flood Statistics in India *[Source: http://floodlist.com/tag/india]*

Year	State Affected	Socio and Economic Loss
2013	Uttarakhand	More than 5700 human causalities
2015	Gujarat	More than 70 human causalities
2015	Andhra Pradesh and Tamil Nadu	More than 500 human causalities
2016	Assam	1.8 Million peoples affected
2017	Gujarat	More than 200 human causalities
2018	Kerala	More than 445 human causalities
2019	Kerala, Madhya Pradesh, Karnataka, Maharashtra and Gujarat	More than 850 human causalities
2020	Assam, Kerala and Bihar	More than 95 human causalities

Statisticsclearlydepictthatflood-basedcausalitiesareunavoidable,butitcouldbereduced by pre-flood measures, through alerts, to evacuate the people. The ultimate reason for flood is heavy rain and in few exceptional cases floods have occurred due to human mistakes and improperdrainagesystem.Floodsduetoheavyrainfallnotonlyaffectstheproperties,butalso affects the day to day life of people.

In general, various sources are attributed to floods but in practice, floods are categorized based on its originating characteristics and risk factors such as fluvial, coastal flood, reservoir flood,sewerflood,burstwaterflood.Riverfloodingorfluvialoccurs,whenthecapacityofthe river exceeds. If the river fails to accommodate the volume of incoming water, then this type of floods occurs. Coastal flood occurs due to the storm which results in a rapid increase of sea level.Theeffectsofcoastalfloodsarehighduetohightidesandsurges.Reservoirfloodsoccur inartificiallakessuchasdams.Duetostructuraldamage,waterleakageinthereservoirdueto erosion or damage leads into reservoir floods, which clears the properties in nearby locations. Floods due to overflow in sewerage systems are termed as sewer floods. It occurs on heavy rainfall due to which the sewerage system is not able to handle the rainwater. This arises due totheheavypressureandcloggedwastematerials.Burstwaterfloodisararefloodingtype which occurs due to major breakage in water supply system. Due to this flood, the chances of human casualties are less, but it affects the transport system in that region.

Technology developments in information systems helps to reduce the huge losses due to floods,bysensingthegeographicalinformationsuchassoiltexture,landscapeetc.,alongwith rainfallratio,satelliteimagesandadditionalradarinformation,whichhelpstopredicttheflood occurrence., As a result, the human casualty rate could be avoided or reduced. Generally, congestedareasareheavilyaffectedduetofloods,whichreducesthelifespanofpeople.Hence, safetyisanessentialfactorinfloodingtimes.Safetycouldbeensuredbycollectingproperdata to analyse the particular region through specific parameters such as wetness index, land slope angle, stream power, stream density, rainfall, curvature and distance. The flood crossing limit is generally termed weighted standardized risk factor, which is used to specify the floodlimits consideringtherapidfloodoccurrenceinaregion.Insomecases,floodsmayoccurduetopoor drainage system and these kinds of floods are termed as manmade floods. In order to analyse the flood nature, proper analysis of drainage system is essential, which describes the stream behaviour. The relation between the streams are need to be analysed as remote sensing data through geographical information system, which helps to gather ~~obtain~~ the information about the region, where the floods might occur. Floods in crowded regions produce huge losses due to the physical models. Without rainfall prediction and flood information, the collateral damage in crowded regions are huge innumbers.

Prediction of rapid floods could be possible in recent times due to the integration geographical information system and remote sensing information. This integrated module could provide the flood occurrence based on the database and geological maps by processing the data efficiently. As a result, the exact location could be predicted. Similarly, Satellite and Radar images help to predict the flood through proper analysis of maps and its topographical information. Soil moisture is used as another factor to determine the flood probability. With the help of remote sensing processes the information about the flood could be obtained through surface soil moisture. The data could be collected~~obtained~~ through advanced microwave scanning radiometer, which measures the temperature facts as C and Xbrightness.

Similar to flood prediction analysis, flood risk assessment is essential, which describesthe flood intensity and its risk factors based on the interaction between depth and areal dimension offlood.Socio-economicelementswhichareaffectedduetofloodscouldbeanalyzedthrough flood risk assessment systems. It is related by anequation

□(□)=□□□□.□□□□□□□□□□□□□□□□□□□□×□□□□□□□□□□□□□□□ (1)

Flood risk assessment is used to identify the root cause for the risks, determine the flood risks and to provide proper guidelines before and during floods. Based on the above flood risk assessment steps is summarized and depicted in figure 1.

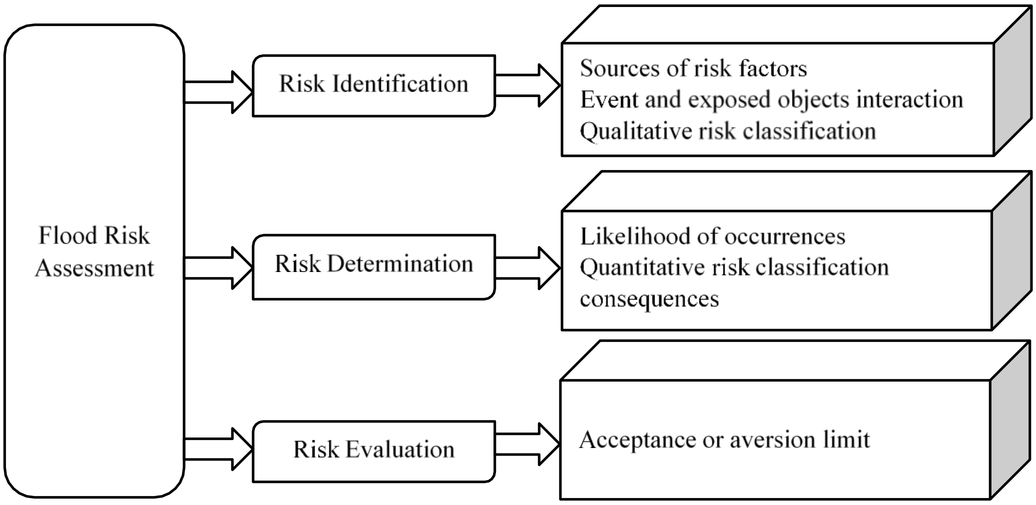


Figure 1. Flood Risk Assessment

Flood prediction is significant at macro level, since, it has vital contribution in prediction and management of water resources, which helps the people to get updates about weather and rainfall status. Through proper prediction and risk assessments, the human lives couldbesavedandthemassivedamagetopropertiescouldbereducedbyredirectingthewater flow using suitable prevention methods. Accurate prediction helps to alert the authorities so thatupcomingdestructionscouldbeminimized.Floodpredictionimpressivelyhelpstomanage water scarcity by redirecting the water to suitable reservoirs, which would be useful in future. Based on the flood prediction mechanism, people could manage and plan the activities to safeguardtheirlivesandbelongingswhichreducesthehugelosses.But,duetothecomplexity, floodpredictioninaparticularregionisatedioustask.Hence,itisessentialtodevelopamodel to obtain a solution, which reduces the flood damages through mapping-based detection of flood prone region. Furthermore, risk assessment models need to map the sensitivity region to avoid huge potential damage. Considering these basic facts, the proposed flood prediction and riskassessmentisdevelopedinthisresearchworktoreducethehumaneffortswithitsaccurate prediction process and thereby save more human lives and properties. The research work is organizedasfollows:sectionIIprovidesin-depthanalysisofexistingfloodpredictionmodels.

Proposed machine learning approach is presented in section III, experimental results and its discussion is given in section IV, followed by conclusion in the last section.

II. RELATEDWORK

The importance of flood prediction and risk assessment has been discussed through various approaches. A vast survey has been made to analyse the methodologies, merits and demerits, which helps to improve the proposed flood prediction and risk assessment model. In general, flood risk assessment and frequency of flood is a stationary assumption process, but the flood probabilities change every year due to the climate changes. Prior knowledge of the climatechangeshelpstopredictthemagnitudeoffloods.Basedonstreamflow,nonstationary risk in floods across Negro river in Brazil is analysed [1] based on the extreme value distribution, such as sea surface temperature, at beginning of the year and previous year data, along with nonstationary parameters, which help to predict the annual peak in the river. This could help the people by providing an early alert under dynamic risk assessment process. This prediction model is limited to specify the river floods and not suitable for other floods.Highly populatedareasundergosevereissuesduetoflooding.Theseurbanfloodshavetobepredicted withhighspatiotemporalresolution,whichisquiteacriticaltask[2].Similarly,earlywarnings for the specified region with cost effective measures is essential in urban locations. Based on the spatiotemporal data, this early prediction and alerts could be provided to the people in highly populated area with high accuracy and specificity which greatly reduces the simulation error and its effects.

Therisktowardsfloodedhomesanditsrestorationprocess[3]isimportantasitdirectly affects the people and produces unnecessary delays in case of restoration. Risk management helps to understand the behaviour of flooded buildings and its contaminants within the building,afterfloods.Anintegrateddryingmodelusingsimulationlinkedmicrobialmodeland geolocationinformationsystemhelpsthepeopletorecoverthebuildingfromthecontaminants. Thisprocesswillhelpthehealthauthorities,localgovernmentsandinsurancecompaniestobe prepared for urban flood risks. In order to assess the floods based on water level and land elevation[4],ahydrodynamicmodelwithdynamiclandflow,floodwallsanddrainagesystem helps to estimate the wave heights and nearshore effects of floods. Lack of backwater precautionmeasureresultsintostormindrainagesimilartoovertoppingfloodsinnaturaldams. The accuracy of coastal flood prediction [5] could be achieved through hydrodynamic model usingtwo-dimensionalhydrodynamicfloodmodel.Thetopographicandtidaldataareanalyzed through maximum likelihood classification as a supervised approach which differentiates the land cover classes. Hydrodynamic models are reliable and provide better statistics. Subsequently,urbanbackshorefloodscouldbehandledwithprecautionmeasureswhichcould save humanlives.

Overtopping of dam is one of the reasons for floods, compared to flood prediction due to heavy rain. This type of prediction process is simple and more accurate. Analytical identificationofparametersfordamfailureneedstobeperformedtopredictthepeakdischarge and outflows. Using model calibration [6], the discharge hydrograph is observed and flood nature is predicted, which helps to reduce the difficulties in risk management process, on and afterflooding.Similarly,observingthewaterdischargepropertiesinadamhelpstopredictthe floods.Sincethetransferfunctionisrelatedtoaverageamountofrainfallandwaterdischarge, flood prediction could be reduced to a simple solution [7], facilitating early warnings which reduces the probability of humancasualties.

Relative analysis of rainfall and runoff in flood prediction is a challenging task. In literature, the predictive capabilities are investigated through Xinanjiang model [8] which is a familiar flood forecasting simulation process in China. Based on geomorphological instances, hydrograph and Xinanjiang model based improved flood prediction is attained in the research model. The limitation of this hybrid model is its hourly time scale which requires catchments causing an increase in the system complexity. Rapid prediction system has ability to represent the rainfall and runoff as risk analysis. This system requires accurate storm surge data due to the computational demand and hence a strong high-fidelity hydrodynamic mode is required, with advanced circulation to represent the rainfall and runoff [9]. This statistical model provides the relationship between the pressure field and wind characteristics, which helps to determine the peak flood responses and flood risk probabilities.

In flood forecasting, collective forecasts are essential in order to obtain results through numerical weather prediction model [10]. Weather prediction provides details about rainfall forecasts based on the hydrological model, which could reduce the after-flood effects by prior warnings to people. Numerical weather prediction requires numerous attributes to obtain the forecast results, which is observed to a significant limitation of the research model. The location and natural features of river faces flood risks due to change in climate. In order to investigate the risk in this region, one dimensional hydrodynamic model is suggested [11] as real time operations of pumps and sluices. Using dual lumped hydrological model, the rainfall and runoff process is estimated for rural and urban areas. In order to avoid flood hazards, the difficulties in draining is considered through evaluation parameters, which reduces the flood severity in river management. Computation time, accuracy level and robustness are the major issuesintraditionalfloodpredictionprocess.Onedimensionalanalysishasevolvedtoimprove the parameters, though the performance lags under critical circumstances. Following this, two dimensional models [12] have emerged to be better flood prediction models. These two-dimensional water models are simulated with GPU to recreate the real time floods. The numericalinstabilitiesarecontrolledthroughfirstorderfinitevolumeprocesswhichissuitable for complex landscape analysis. Bayesian network [13] based models are further evolved to reduce the uncertainties in two dimensional models which provides better risk analysis for floods inreservoirs.

In some cases, the flood stage reliability and its predictions get affected due to the poor performance of hydraulic model uncertainties. Due to these uncertainties, prediction through conventional hydraulic model becomes inefficient. In order to improve the prediction performance, Bayesian model averaging [14] is used to combine the prediction results from different hydraulic model, which provides better stage prediction and distribution. Integrating thechannelshape,width,inputflowandroughness,theuncertaintiesarereducedandvalidated the floods. Hierarchical Bayesian model [15] is developed for rapid flood analysis in ruraland urban places of China. Based on the annual daily flow obtained from 15 different stream flow gauges sites, the proposed approach analyses the generalized distribution value and flood distribution. To obtain the posterior distribution, Markov chain Monte Carlo method is used with Gibbs sampling. Compared to conventional delta model, the proposed approach shows better performance credibility for flood quantiles. Moreover, the proposed approach allows regional analysis, as index flood method, to obtain region homogeneity and scaleinvariance.

Multi criteria decision making analysis techniques and Machine Learning models are testedtoobtainthefloodsusceptibilityofChinainliterature[16].Inordertoobtainthedesired results,inputparameterssuchasnormalizedDifferenceVegetationIndex,distancefromriver, curvature, Lithology, land use, altitude, Topographic Wetness Index, Stream Transport Index, slope, Stream Power Index, soil type and rainfall are considered for analysis. Amongst these, NBT approach obtains better flood prediction compared to other conventional prediction models. Same input parameters, basins and flood susceptibility are reported in literature [17] through analytical hierarchical process, knowledge driven, fuzzy logic and logistic regression models. Comparative analysis of research work provides better area under curve analysisover

Dwarkeswar river. Multi variant discriminate analysis [18] using regression trees and support vectormachinesareevolvedtoobtainbetterfloodpredictionanditssusceptibilities.Compared to multi criteria-based decision analysis, these discriminate analyses provide accurate results forurbanfloods.Similarly,multi-layercapacitanceresistivemodelsareusedtoobtaintheflood prediction in reservoir. Considering each layer, the injection and production rates constituted by individual and combined, the bottom hole pressure could be obtained. Kalman filter [19] is used to estimate the coefficients of each layer and the indexnumbers.

Flood prediction using Machine Learning [20] and Artificial Intelligence [21] approaches is the recent research trend which predicts the floods based on sensor data and spatiotemporal data of floods. However, researchers are working on it to provide a better predictionmodel.Integratedlongshort-termmemoryandreducedordermodelisusedtoframe the prediction model in literature [22]. The proposed model has ability to represent the distribution of floods with reduced spatial datasets, which is obtained through single value decomposition and orthogonal decomposition. These reduced dataset increases the efficiency in terms of accuracy and computation time. Neural network-based flood prediction isreported

[23]toimprovethepredictionaccuracywithminimizedcomputationtime.Integratedorhybrid approaches are evolved to improve the performance of machine learning models. Resampling algorithm reported in literature [24] is combined with machine learning model provides better classification and predictionperformance.

Urban flood data is either structured or mostly unstructured in nature. Based on this data, an urban flood database is created to predict the urban flood depth using deep learning algorithm [25] such as gradient boosting decision tree. Rainfall and its duration, peak evaporation and utilization of lands for roads, water bodies, properties and grasslands are consideredalongwithrainfallperiodstopredictthefloodmap.Regressionmodelisusedalong with proposed approach to obtain the conditioning factors. From the above survey, it is observed that precise flood prediction and risk assessment is still lacking. Formulation of the flood prediction based on previous and current data similarities will provide better prediction and assessment model with high reliability and accuracy. Since inaccurate prediction accrues huge property loss and waste of time, such models lead into an inefficient risk management. Therefore,itisessentialtoformulatethefloodpredictionsystemalongwithriskassessmentas a precise model compared to existing floodpredictors.

III. PROPOSEDWORK

Mathematical model for risk assessment from big data through hybrid differential evaluationandadaptiveneurofuzzyinferencesystemispresentedinthissection.ANFISisan advanced neuro fuzzy based reliable estimator, which provides solutions to complex issues. Combining the artificial neural network and fuzzy logic, ANFIS provides high learning capability. The proposed hybrid model includes advanced fuzzy rules to define the nonlinear functions in flood model. Through artificial neural network learning rules, the parameters are identified and structured for fuzzy inference system. Using the given dataset and neural networktrainingprocess,ANFISmodelidentifiesthemissingrulessothatitprovidesaccurate learning, strong implementation and reliable results. Flood modelling through ANFIS helps to analyseshorttermforecastsofrainfallandfloodswithdifferenttypesofdataresultsintoahigh accuracy prediction system. The proposed model is designed with differential evolution algorithm coupled with ANFIS to obtain optimized solution. Differential evolution is a population based evolutionary optimization algorithm incorporated with ANFIS in the proposedwork.

The collected big data is used as input to the ANFIS model and the results are obtained as a prediction model. In order to improve the training process, the dimensionality of the data needstobereducedalongwithimprovedfeatureselection.Toattainthis,differentialevolution model is incorporated with ANFIS model to provide solution to the complex data dimensions. Fourprocessofdatareductionandselectionisperformedthroughdifferentialevolutionmodel such as initiation, mutation, crossover and selection.

In the initiation process, the initial population is randomly generated. Based on the collected information, the initial features are selected randomly with specified design parameters, so that each value is obtained as

$$\square_{\square,} = \square_j^{\square\square\square} + \square_{\square}\left(\square_j^{\square\square\square} - \square_j^{\square\square\square}\right) \qquad (2)$$

where$\square = 1,2 \ldots \square_{\square}and\square = 1,2 \ldots \square_{\square}$whichisadecisionvariable.Theupperandlower boundsofdatadimensionsarerepresentedas$\square_{\square}^{\square\square\square}and\square_{\square}^{\square\square\square}$respectivelywhichliesinthe range of [0,1]. Mutation is the second process in differential evolution in which a mutation operator generates differential vectors for the given input data vectors and it is give as

□□,□+1=□□1,□□□ +□(□□2,□□□−□□3,□□□) (3)

where randomly selected indices are given as □1, □2and □3, randomly selected population are given as □□1,□□□ , □□2,□□□ and □□3,□□□ . The scale factor is represented as □. Using target and donor vectors a trail vector is produced in the third stage of differential evolution process i.e., cross over process and it is given as

□□,,□□+1 □□ □□ ≤ □□ □□ □ =□

□′

□,,□□+1 = { □□□□ (4)

□□,□□□

where the uniform random number is given as □□which lies in the range of [0,1], similarly the crossover function is given as □□which is in the range of [0,1]. The selection of minimum parameter and the mutant vector is given as □. Selection is the last process in differential evolution where the fitness function of trail vector and the target vector is compared. Based on the obtained values, the best choice for next generation is selected in the selectionprocess.ThishelpstoobtainthenecessaryfeaturestobegivenintoANFISmodelso that the accuracy of the prediction system could be improved. The selection process is given as

□′□,□+1 □□□(□′□,□□□+1)≤□(□□,□□□)

□□,□+1 = { (5)

□,□□□ □□ □□

where the parent vector is represented as □□,□□□and the fitness function is given as □(□□,□□□).

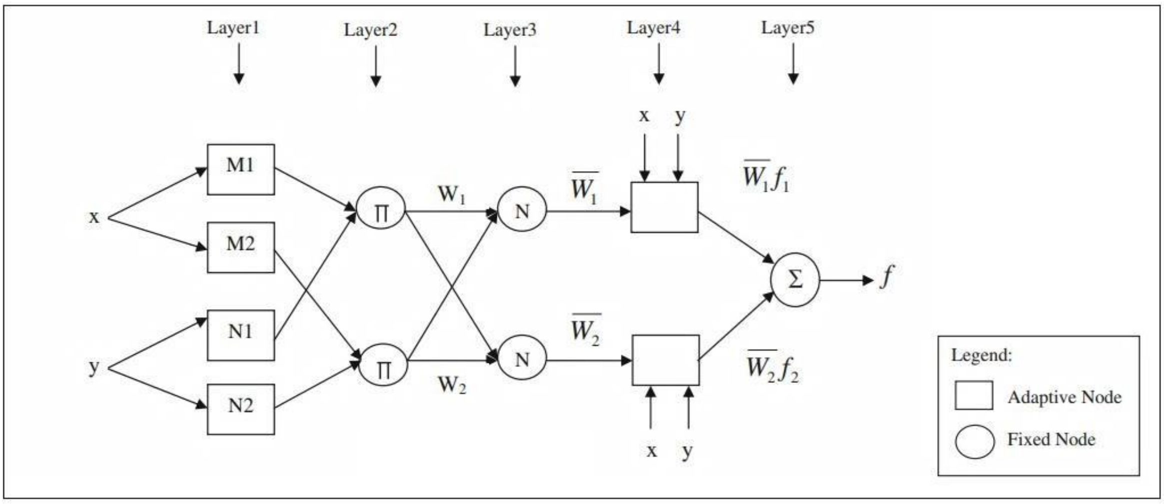


Figure 2. ANFIS System

The proposed system is a Sugeno Fuzzy system based on feed forwarding network design which is typically included in the neural network model. It provides better adaption ratio due to its characteristics. Figure 2 depicts the two input ANFIS model with one output. Based on the input conditions the rule for ANFIS is derived as

If□is$□_1$and□is$□_1$then□=$□_1$□+$□_1$□+$□_1$ (6)

If□is$□_2$and□is$□_2$then□=$□_2$□+$□_2$□+$□_2$ (7)

where□ and □ are the inputs and □ is the output of the system with in the region specifiedbytherules, $□_1$,$□_1$,$□_2$,$□_2$arefuzzysetsandthedesignparametersaregivenas $□_1$,$□_1$,$□_1$,$□_2$,$□_2$,$□_2$thesedesignparametersareobtainedatthetimeoftrainingprocess.Fuzzy set is used to represent the node in the layer and output layer and its corresponding layers are related to membership degree function of fuzzy sets. The shape parameters are used to define themembershipfunctionandabell-shapedfunctioniswidelyusedinANFISmodel.Thesame is used in the proposed work. The membership function is represented as $□_{□□}$ and it is givenas

$□_{□□}$ = 1 / (1+((□−□□)/□□)^{2□□}) (8)

where□ is the input value for $□^{□h}$ node, the membership functions and its parameters are representedas$□_□$,$□_□$,$□_□$andthisparametersareconsideredasconditionalparametersofthefirst layer. Based on the degree of activation of rules, every node in the second layer computes the input and performs a multiplication process. The multiplied membership function is givenas

$□_□$=$□_{□□}$(□)$□_{□□}$(□) for □=1,2... (9)

where $□_{□□}$ (□) is the membership function of $□_□$rule set for the value □ and $□_{□□}$ (□) is the membership function of $□_□$rule set for the value □. In the third layer, the ratio of $□^{□h}$ rule activitydegreetothesumofallactivitydegreeiscomputedanditisrepresentedas$□_□$which is a normalized degree for $□^{□h}$ rule and it is givenas

$□_□$ = □□ / (□1+□2) for□=1,2... (10)

Fourth layer provides the output of any layer and it is obtained as

$$\bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i) \tag{11}$$

wheretheconsequentparametersaregivenas□,□□□□□whicharechangeable.Inthe fifth layer, the final output of the network is obtained as function □ which is a summation incoming signals from all the nodes. Generally, the total number of nodes is similar to output parameter which is derived in the fifth layer is givenas

$$\sum \square\square \; = \frac{\sum \square \; \square\square\square}{\sum \square\square} \qquad (12)$$

□ □

In order to obtain fast training results, the network parameters are adjusted in the network by keeping fixed premise parameters and a forward propagation is considered for network to layer information transfer. Similarly, in backward phase, the selected parameters are tuned into fixed values and error propagation is initiated. Using gradient descent method, these premise parameters are modified. The membership function is the only parameter which needs to be specified by the user in the learning algorithm.

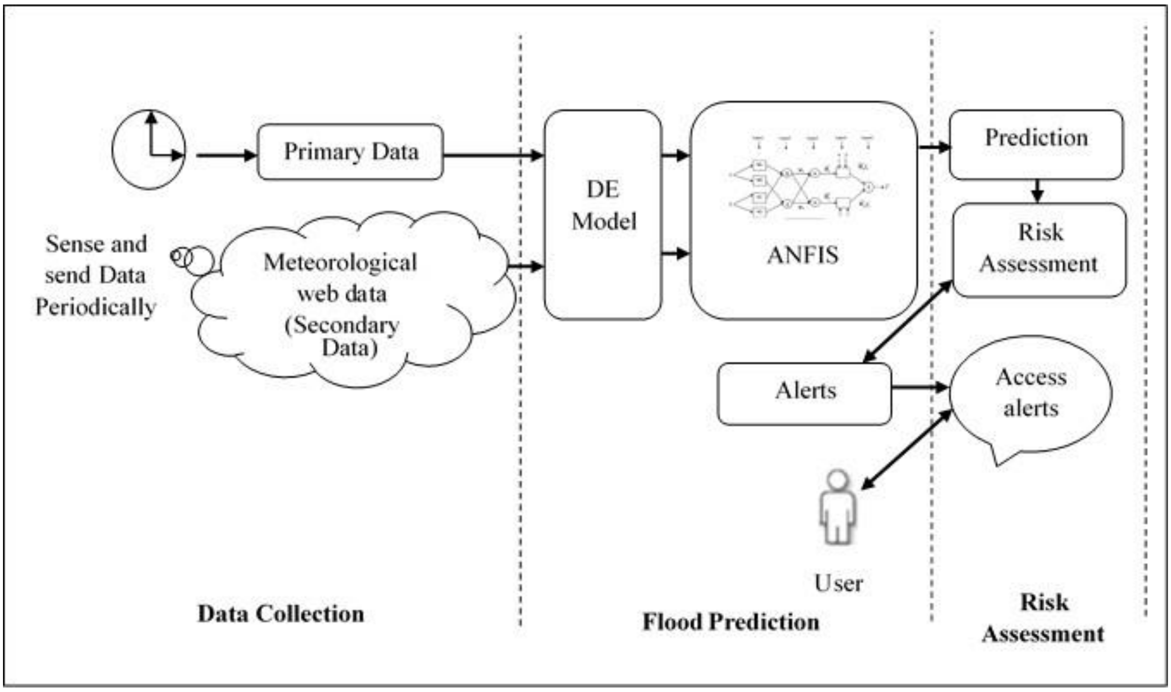


Figure 3 Proposed Flood Prediction and Risk assessment model

Figure 3 depicts the process flow of proposed flood prediction and risk assessment model. Based on the periodic sensing of primary and secondary data from real time sensors andmeteorologicaldata,theinformationabouttherainfallstatisticsisobtained.Thiscollected informationconsistsofstructureddatafrommeteorologicalandunstructureddatafromsensors which needs to be pre-processed. Using differential evolution model, these unstructured data are processed and reduced with equal dimensions similar to meteorological data. Theselected

features are fed into ANFIS structure to obtain desired classification results. Based on the predictionresults,riskassessmentisperformedbyprovidingalertstotheusersintheparticular region. Based on proposed approach, the early prediction of flood could be possible and throughriskassessmenttheusergetsclearideaabouttherainfallstatusandfloodprobabilities.
Summary of the proposed work pseudocode is given asfollows.

Input: Sensor data, Meteorological data

Output: Flood Prediction and risk assessment

Initialize serial communication with host

Read the primary data

Send data to system

Read the meteorological data

Send data to system

Generate Initial Population of size Np
While

For each individual, j is the population

Generaterandomintegers□□,□□and□□withrespecttoinitialpopulation
Generate random integer igen ∈ (1,n)

For each parameter of *i*

□′□,□□□+□

□□□(□′□,□□□+□)≤□(□□,□□□)

□□,□□□+□={□ □□□□

□,□□□

End for

Replace □□with child □′□, if □′□is better
End

Train 75% of data base
Validate 25% of database

Test the vector rules☐☐

Test the initial learning rates ☐☐☐

Set number of test values
Set number of epochs

For ☐ ∊ ☐☐☐do For ☐ ∊ ☐☐do

For ☐ = ☐: ☐ do

Generate random FIS with 2 inputs and learning rate ☐ = ☐

EvaluateANFISfunctionwith☐=☐☐,trainingandvalidationofrandomgeneratedFIS
Save the FIS with lowest validation error, training error and vector validation

End for
End for

End for

IV. RESULTS ANDDISCUSSION

The proposed model is experimented by analysing the rainfall status of the Gujarat state whichislocatedonWesterncoastofIndianPeninsula.Aspertherecentsurvey,Gujaratisthe fifth largest state by area and ninth largest state by population where 60.4 million people live across the state. The largest coast line of 1,600 km will provide seasonal rainfall to the state. The average humidity will be 67% and the average rainfall will be 1,107 578 mm. Figure 4 depicts the topographical view of Gujaratprovince.

Figure 4 Topographical view of Gujarat province.

The average rainfall for the past ten years in Gujarat is depicted in figure 5. The rainfall data collected from GSDMA for the years 2010 to 2019.

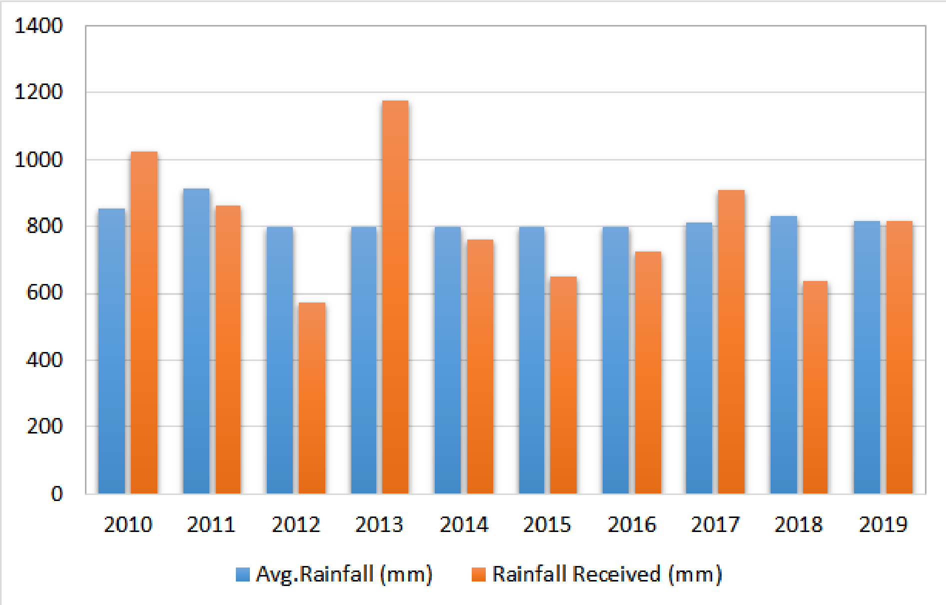


Figure 5. Average rainfall for a decade in Gujarat

The annual rainfall for Gujarat state is depicted in figure 6. Gujarat faced more than 100% of rainfall in three years which is considered for flood prediction and risk assessments.

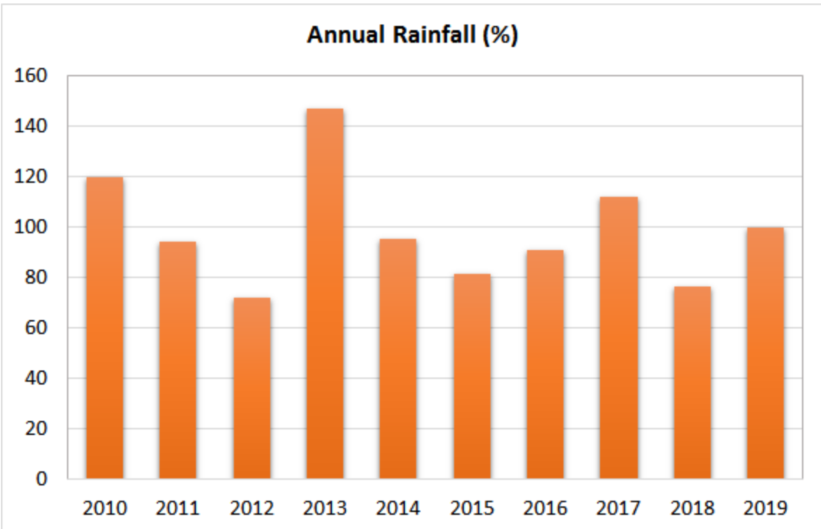
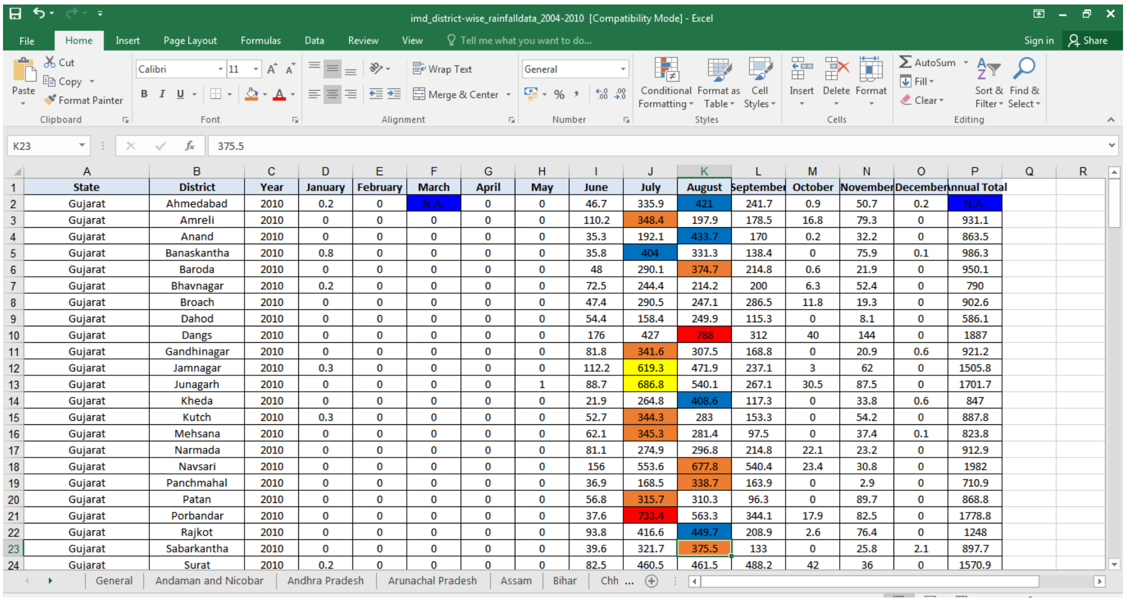


Figure 6. Annual Rainfall Statistics in Gujarat

It is observed from the average rainfall and annual rainfall statistics, there is minor variation in every year rainfall cycle but it has a constant rain circle in the decade. Therainfall circle of the monsoon season ~~circle~~startsfromJulyandendsinAugustorSeptember, where most of the regions are hot, wet and more humid. Similarly, the high distribution of monsoon rain indicates that the summer season is hot with almost zero mm of rain.



State	District	Year	January	February	March	April	May	June	July	August	September	October	November	December	Annual Total
Gujarat	Ahmedabad	2010	0.2	0	N.A.	0	0	46.7	335.9	421	241.7	0.9	50.7	0.2	N.A.
Gujarat	Amreli	2010	0	0	0	0	0	110.2	348.4	197.9	178.5	16.8	79.3	0	931.1
Gujarat	Anand	2010	0	0	0	0	0	35.3	192.1	433.7	170	0.2	32.2	0	863.5
Gujarat	Banaskantha	2010	0.8	0	0	0	0	35.8	404	331.3	138.4	0	75.9	0.1	986.3
Gujarat	Baroda	2010	0	0	0	0	0	48	290.1	374.7	214.8	0.6	21.9	0	950.1
Gujarat	Bhavnagar	2010	0.2	0	0	0	0	72.5	244.4	214.2	200	6.3	52.4	0	790
Gujarat	Broach	2010	0	0	0	0	0	47.4	290.5	247.1	286.5	11.8	19.3	0	902.6
Gujarat	Dahod	2010	0	0	0	0	0	54.4	158.4	249.9	115.3	0	8.1	0	586.1
Gujarat	Dangs	2010	0	0	0	0	0	176	427	788	312	40	144	0	1887
Gujarat	Gandhinagar	2010	0	0	0	0	0	81.8	341.6	307.5	168.8	0	20.9	0.6	921.2
Gujarat	Jamnagar	2010	0.3	0	0	0	0	112.2	619.3	471.9	237.1	3	62	0	1505.8
Gujarat	Junagarh	2010	0	0	0	0	1	88.7	686.8	540.1	267.1	30.5	87.5	0	1701.7
Gujarat	Kheda	2010	0	0	0	0	0	21.9	264.8	408.6	117.3	0	33.8	0.6	847
Gujarat	Kutch	2010	0.3	0	0	0	0	52.7	344.3	283	153.3	0	54.2	0	887.8
Gujarat	Mehsana	2010	0	0	0	0	0	62.1	345.3	281.4	97.5	0	37.4	0.1	823.8
Gujarat	Narmada	2010	0	0	0	0	0	81.1	274.9	296.8	214.8	22.1	23.2	0	912.9
Gujarat	Navsari	2010	0	0	0	0	0	156	553.6	677.8	540.4	23.4	30.8	0	1982
Gujarat	Panchmahal	2010	0	0	0	0	0	36.9	168.5	338.7	163.9	0	2.9	0	710.9
Gujarat	Patan	2010	0	0	0	0	0	56.8	315.7	310.3	96.3	0	89.7	0	868.8
Gujarat	Porbandar	2010	0	0	0	0	0	37.6	733.4	563.3	344.1	17.9	82.5	0	1778.8
Gujarat	Rajkot	2010	0	0	0	0	0	93.8	416.6	449.7	208.9	2.6	76.4	0	1248
Gujarat	Sabarkantha	2010	0	0	0	0	0	39.6	321.7	375.5	133	0	25.8	2.1	897.7
Gujarat	Surat	2010	0.2	0	0	0	0	82.5	460.5	461.5	488.2	42	36	0	1570.9

Figure 7 Rainfall data of Gujarat *[Source: https://www.indiawaterportal.org/]*

For experimental analysis, rainfall data of Gujarat is considered and the simulation is performed based on the big data from various sources like humidity, temperature, soil characteristics etc., Experimentation is performed in python, with the above-mentioned parameters for predicting the average rainfall during the months of June to August. Based on the rainfall data and simulation results, the performance of proposed model is evaluated. Correlation between predicted rainfall and measured rainfall is depicted in figure 8.

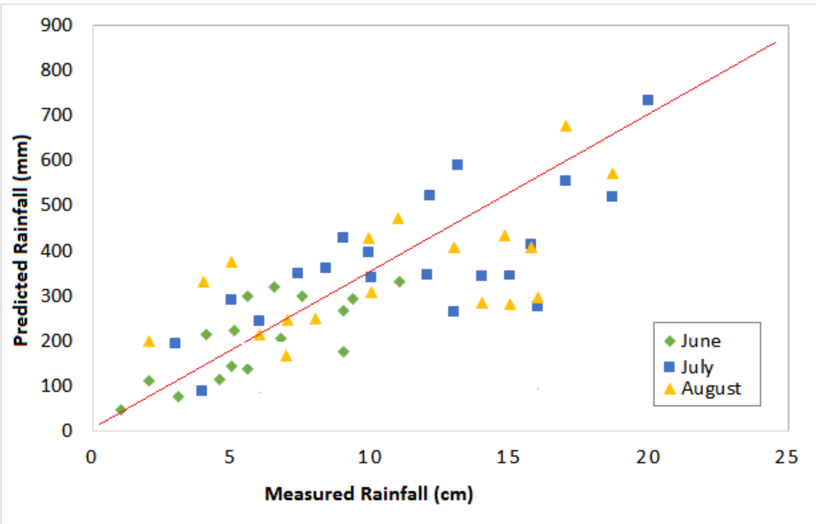


Figure 8. Correlation between predicted rainfall and measured rainfall

The correlation between the predicted flow and measured flow ~~and predicted flow~~ is depicted in figure 9. It is observed that the simulation result has small deviations compared to actual data in fewplaces, apart from that the prediction results are near to the actual measuredvalues.

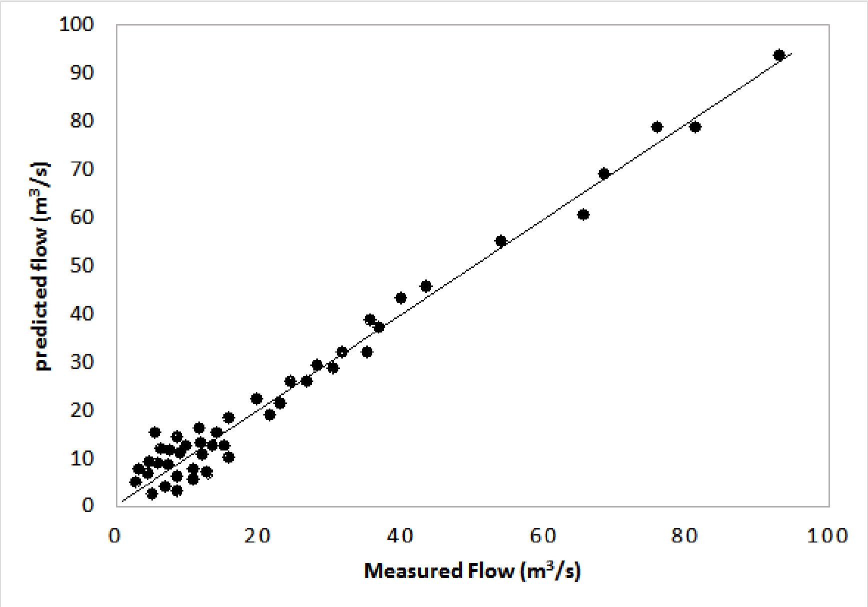


Figure 9. Correlation between predicted flow and measured flow

Performance evolution of ANFIS model on training and validation period is given in table 2.BasedonthefactorslikeRMSE,CorrelationCo-efficient,Discrepancyrationstheperformance isevaluated.

Table 2. Performance evaluation of Proposed Hybrid Model

Ratio: 75-25%			
	RMSE	Correlation Co-efficient	Discrepancy Ratio
Training Phase	550.24	0.99	0.99
Validation Phase	524.22	0.98	0.96

Basedonthefactorssuchaswetnessindex,landslopeangle,streampower,streamdensity, rainfall, curvature and distance, the risk assessment is performed in the proposed model. The parameters used for risk assessment is listed in table 3. The predicted values are analyzed and observedintermsofriskassessmentparametersandthenecessaryactionsaretakentoalertthe public.

Table 3. Parameters for Risk Assessment

S.No	Parameter	Range
1	Wetness index	0.4-0.6
2	Land Slope angle	8°-15°
3	Stream Power	30-40kmph
4	Stream Density	Average density of 290 per sq. km
5	Rainfall	1,107 578 mm
6	Curvature	Concave, convex and flat
7	Distance	More than 50km

Considering the above parameters, the risk assessment performance of proposed model is observed and compared with other conventional models. The performances of conventional models are satisfactory but due to reduced data dimension and learning rates, proposed hybrid model attains a noteworthy performance in risk assessment. Figure 10 depicts the risk assessment comparison of all the three models.

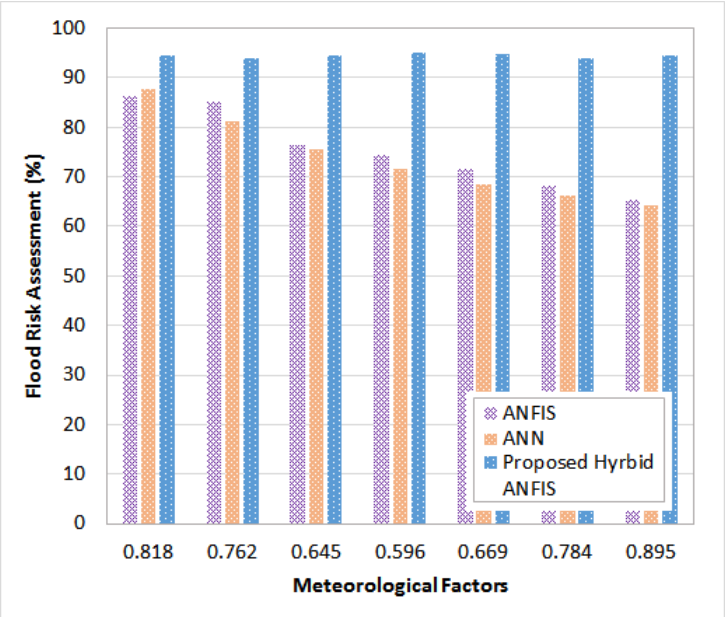


Figure 10. Risk assessment performance comparison

Performance comparison of proposed hybrid model and conventional ANFIS andartificial neuralnetworkisdepictedinfigure10intermsofprecision,recallandf-measurescores.Table 4 gives the comparative analysisdetails.

Table 4. Performance Comparison

Model	Precision	Recall	F-measure
ANFIS	0.91	0.90	0.42
ANN	0.93	0.89	0.50
Proposed Hybrid ANFIS	0.96	0.93	0.64

From above table 4, it is observed that proposed hybrid model attains better performance compared to conventional models due to its data handling capabilities.

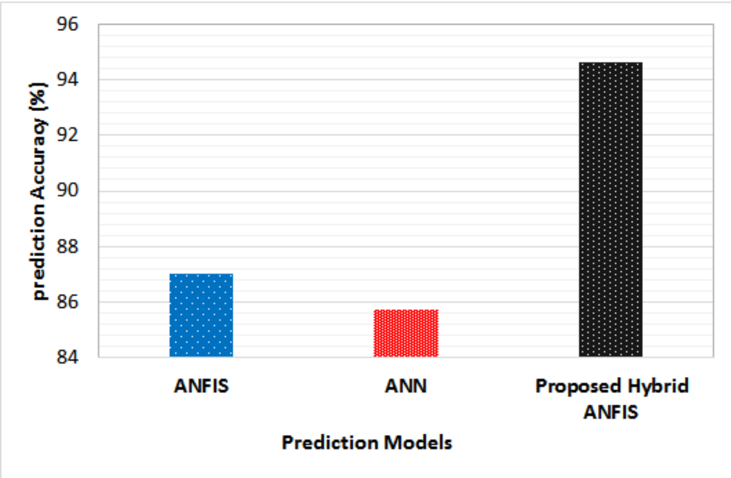


Figure 11 Accuracy comparison

Flood prediction accuracy of all the three models are compared and depicted in figure 11. Based on the historical data analysis, the proposed hybrid approach provides better performance over conventional ANFIS and artificial neural network model.

V. CONCLUSION

A novel flood risk assessment model is proposed in this research work using differential evolution and adaptive neuro fuzzy inference system as a hybrid approach. The importance of flood prediction and risk assessment process is presented in the article to obtain prior knowledge about the flood-oriented research issues. The dimensionality issues in big data is reduced using differential evolution model and then, the selected feature data vectors are fed into the adaptive neuro fuzzy inference system to obtain the flood prediction results. Based on thepredictionresultsandtheriskassessmentparameters,thenecessaryproceduresareinitiated asriskmanagementprocess.Experimentalanalysisofproposedworkiscarriedoutwith historical data and validated with experimental results. Compared to conventional adaptive neuro fuzzy inference system and artificial neural network model, proposed hybrid model attains better prediction accuracy. The relative analysis experimented in this research work helps to predict the future floods, which could save human lives. The limitation of the model is in its data sample as it considers the monthly statistics and hence could not provide daily predictions. Accuracy of the system may slightly change due to global impacts and climatic changes. The prediction performance could be further improved in future using hybridmodels with other differentparameters.

REFERENCES

1. Carlos H.R. Lima, UpmanuLall, Tara J. Troy, NareshDevineni (2015), "A climate informed model for nonstationary flood risk prediction: Application to Negro River at Manaus, Amazonia" Journal of Hydrology,522:594-602.
2. ArezooRafieeinasab, Amir Norouzi, Sunghee Kim, HamidehHabibi, BehzadNazari, Dong-JunSeo,HaksuLee,BrianCosgrove,ZhengtaoCui(2015),"Towardhigh-resolution flash flood prediction in large urban areas – Analysis of sensitivity to spatiotemporal resolution of rainfall input and hydrologic modelling" Journal of Hydrology,531(2):370-388.
3. JonathonTaylor, Ka man Lai, Mike Davies, David Clifton, Ian Ridley, Phillip Biddulph (2011),"Floodmanagement:Predictionofmicrobialcontaminationinlarge-scalefloodsin urban environments", Environment International,37(5):1019-1029.
4. Timu W. Gallien, Brett F. Sanders, Reinhard E. Flick (2014), "Urban coastal flood prediction: Integrating wave overtopping, flood defences and drainage", Coastal Engineering,91:1-28.
5. AvideshSeenath(2015),"Modellingcoastalfloodvulnerability:Doesspatially-distributed friction improve the prediction of flood extent?", Applied Geography,64:97-107.
6. Hongbo Ma, Xudong Fu (2012), "Real time prediction approach for floods caused by failure of natural dams due to overtopping" Advances in Water Resources,35:10-19.
7. YuliantiHasanah, MarizsaHerlina, Hilda Zaikarina (2013), "Flood Prediction using Transfer Function Model of Rainfall and Water Discharge Approach in Katulampa Dam", Procedia EnvironmentalSciences,17:317-326.
8. Cheng Yao, Ke Zhang, Zhongbo Yu, Zhijia Li, Qiaoling Li (2014), "Improving the flood prediction capability of the Xinanjiang model in ungauged nested catchments bycoupling

it with the geomorphologic instantaneous unit hydrograph”, Journal of Hydrology,517:1035-1048.

9. Benjamin Bass, Philip Bedient (2018), “Surrogate modeling of joint flood risk across coastal watersheds”, Journal of Hydrology,558:159-173.

10. Wansik Yu, Eiichi Nakakita, Kwansue Jung (2016), “Flood Forecast and Early Warning with High-Resolution Ensemble Rainfall from Numerical Weather Prediction Model”, Procedia Engineering,154:498-503.

11. Zi-junHu,Ling-lingWang,Hong-wuTang,Xiao-mingQi(2017),“Predictionofthefuture flood severity in plain river network region based on numerical model: A case study”, Journal of Hydrodynamics,29(4):586-595.

12. Isabel Echeverribar, Mario Morales-Hernández, P. Brufau, P. García-Navarro (2019),“2D numerical simulation of unsteady flows for large scale floods prediction in real time” Advances in Water Resources,134:1-16.

13. QingwenLu,Ping-anZhong,BinXu,FeilinZhu,YufeiMa,HanWang,SunyuXu(2020), “Risk analysis for reservoir flood control operation considering two-dimensional uncertainties based on Bayesian network”, Journal of Hydrology,589:1-16.

14. Zhu Liu, VenkateshMerwade (2018), “Accounting for model structure, parameter and input forcing uncertainty in flood inundation modeling using Bayesian model averaging” Journal ofHydrology,565:138-149.

15. Yun-biaoWu,Lian-qingXue,Yuan-hongLiu(2019),“Localandregionalfloodfrequency analysis based on hierarchical Bayesian model in Dongting Lake Basin, China”, Water Science and Engineering,12(4):253-262.

16. KhabatKhosravi, HimanShahabi, Binh Thai Pham, Jan Adamowski, AtaollahShirzadi, BiswajeetPradhan, Jie Dou, Hai-Bang Ly, GyulaGróf, HuuLoc Ho, Haoyuan Hong, Kamran Chapi, Indra Prakash (2019), “A comparative assessment of flood susceptibility modeling using Multi-Criteria Decision-Making Analysis and Machine Learning Methods”, Journal of Hydrology,573:311-323.

17. Sadhan Malik, Subodh Chandra Pal, IndrajitChowdhuri, Rabin Chakrabortty, Paramita Roy, Biswajit Das (2020), “Prediction of highly flood prone areas by GIS based heuristic and statistical model in a monsoon dominated region of Bengal Basin” Remote Sensing Applications: Society and Environment” Remote Sensing Applications: Society and Environment,19:1-16.

18. BahramChoubin, EhsanMoradi, Mohammad Golshan, Jan Adamowski, FarzanehSajedi-Hosseini,AmirMosavi(2019),“Anensemblepredictionoffloodsusceptibilityusing

The Seybold REPORT

ISSN 1533-9211

multivariate discriminant analysis, classification and regression trees, and support vector machines" Science of The Total Environment, 651(2):2087-2096.

19. Zequn Zhang, Heng Li, Dongxiao Zhang (2015), "Water flooding performance prediction by multi-layer capacitance-resistive models combined with the ensemble Kalman filter", Journal of Petroleum Science andEngineering,125:1-19.
20. Snehil,RuchiGoel(2020),FloodDamageAnalysisUsingMachineLearningTechniques", Procedia Computer Science,173:78-85.
21. Romulus Costache,DieuTien Bui (2019), "Spatial prediction of flood potential using new ensembles of bivariate statistics and artificial intelligence: A case study at the Putna river catchment of Romania" Science of The Total Environment,691:1098-1118.
22. Hu.R,FangxinFang,christopherpain,IonelMNavon(2019),"Rapidspatio-temporalflood prediction and uncertainty quantification using a deep learning method" Journal of Hydrology, 575:911-920.
23. Simon Berkhahn, Lothar Fuchs, InsaNeuweiler (2019), "An ensemble neural network model for real-time prediction of urban floods", Journal ofHydrology,575:743-754.
24. EsmaeelDodangeh, BahramChoubin, Ahmad NajafiEigdir, NarjesNabipour, Mehdi Panahi, ShahaboddinShamshirband, Amir Mosavi (2020), Integrated machine learning methods with resampling algorithms for flood susceptibility prediction', Science of The Total Environment,705:1-15.
25. ZeningWu,YihongZhou,HuiliangWang,ZihaoJiang(2020),"Depthpredictionofurban flood under different rainfall return periods based on deep learning and data warehouse" Science of The Total Environment,716:1-15.