VISVESVARAYA TECHNOLOGICAL UNIVERSITY

Belagavi – 590018

A Project Report on

FLOOD PREDICTION USING AI

Submitted by

CHAITANYA THEKKUNJA	**4SN17CS711**
AJAY KARTHIK	**4SN17CS700**
KISHAN KASHYAP	**4SN17CS714**
SHUBHANG	**4SN17CS723**

In partial fulfillment of the requirements for the degree of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE & ENGINEERING

Under the Guidance of

Mrs. GANGOTHRI S

Associate Professor

Department of Computer Science & Engineering

SRINIVAS INSTITUTE OF TECHNOLOGY

(NAAC ACCREDITED)

MANGALURU - 574143, KARNATAKA.

2020 - 2021

SRINIVAS INSTITUTE OF TECHNOLOGY

(NAAC ACCREDITED)

MANGALURU - 574143, KARNATAKA.

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

CERTIFICATE

*Certified that the project work entitled "**FLOOD PREDICTION USING AI**" is a bona fide work carried out by*

CHAITANYA THEKKUNJA	**4SN17CS711**
AJAY KARTHIK	**4SN17CS700**
KISHAN KASHYAP	**4SN17CS714**
SHUBHANG	**4SN17CS723**

*in partial fulfillment for the award of **BACHELOR OF ENGINEERING** in **COMPUTER SCIENCE & ENGINEERING** of the **VISVESVARAYA TECHNOLOGICAL UNIVERSITY, BELAGAVI** during the year 2020 – 2021. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report deposited in the departmental library. The project report has been approved as it satisfies the academic requirements in respect of Project work prescribed for the Bachelor of Engineering Degree.*

Mrs. Gangothri S	**Dr. Shivakumar G. S.**	**Dr. Shrinivasa Mayya D.**
Project Guide	**Head of the Department**	**Principal**

Name of the Examiners **Signature with Date**

1.

2.

ABSTRACT

Floods are among the most destructive natural disasters, which are highly complex to model. The research on the advancement of flood prediction models contributed to risk reduction, policy suggestion, minimization of the loss of human life, and reduction the property damage associated with floods. To mimic the complex mathematical expressions of physical processes of floods, during the past two decades, machine learning (ML) methods contributed highly in the advancement of prediction systems providing better performance and cost-effective solutions. Due to the vast benefits and potential of ML, its popularity dramatically increased among hydrologists. Researchers through introducing novel ML methods and hybridizing of the existing ones aim at discovering more accurate and efficient prediction models.

CHAPTER 1

INTRODUCTION

Flood prediction is the use of forecasted precipitation and stream-flow data in rainfall-runoff and stream-flow routing models to forecast flow rates and water levels for periods ranging from a few hours to days ahead, depending on the size of the watershed or river basin. Flood prediction can also make use of forecasts of precipitation in an attempt to extend the lead-time available. Flood prediction is an important component of flood warning, where the distinction between the two is that the outcome of flood prediction is a set of forecast time-profiles of channel flows or river levels at various locations, while "flood warning" is the task of making use of these forecasts to tell decisions on warnings of floods.

Real-time flood prediction at regional area can be done within seconds by using the technology of artificial neural network. Effective real-time flood prediction models could be useful for early warning and disaster prevention.

1.1 Problem Statement

Among the natural disasters, floods are the most destructive, causing massive damage to human life, infrastructure, agriculture, and the socioeconomic system. Governments, therefore, are under pressure to develop reliable and accurate maps of flood risk areas and further plan for sustainable flood risk management focusing on prevention, protection, and preparedness.

1.2 Existing System

Physically based models were long used to predict hydrological events, such as storm, rainfall/runoff, shallow water condition, hydraulic models of flow, and further global circulation phenomena, including the coupled effects of atmosphere, ocean, and floods. Although physical models showed great capabilities for predicting a diverse range of flooding scenarios, they often require various types of hydro-geomorphological monitoring datasets, requiring intensive computation, which prohibits short-term prediction. Numerous studies suggest that there is a gap in short-term prediction capability of physical models. For instance, on many occasions, such models failed to predict properly.

1.3 Proposed System

This system aims to collect data from all the states of India and form a generalized dataset. A machine learning algorithm is applied to the labelled dataset, and patterns are extracted, which, in turn, obtain maximum accuracy with real-time input. In general, the dataset collected for predicting is split into a Training set and Test set. Generally, 7:3 ratios are applied to split the Training set and Test set. The Data Model is then created using a Multi-Layer Perceptron Classifier, and the resulting data set is then passed through it for prediction.

1.4 Objective

The objective of Flood Prediction using AI is to design a incremental model to predict floods based on rainfall levels, channel flows, river levels etc. An approach of prediction is using Artificial Neural Networks that has very good working efficiency produces the accurate results. The system helps to improve the performance. This method can be used to predict both long term and short term floods in a given region. Maintaining the project is easy and manageable.

CHAPTER 2

LITERATURE SURVEY

[1] Developing mathematical models using Artificial Neural Networks

Artificial Neural Networks (ANNs) are efficient mathematical modeling systems with efficient parallel processing, enabling them to mimic the biological neural network using inter-connected neuron units. Among all ML methods, ANNs are the most popular learning algorithms, known to be versatile and efficient in modeling complex flood processes with a high fault tolerance and accurate approximation. In comparison to traditional statistical models, the ANN approach was used for prediction with greater accuracy. ANNs were already successfully used for numerous flood prediction applications, e.g., stream-flow forecasting, river flow, rainfall–runoff, precipitation–runoff modeling etc. Despite the advantages of ANNs, there are a number drawbacks associated with using ANNs in flood modeling, e.g., network architecture, data handling, and physical interpretation of the modeled system. A major drawback when using ANNs is the relatively low accuracy, the urge to iterate parameter tuning, and the slow response to gradient-based learning processes. Artificial neural networks require processors with parallel processing power, in accordance with their structure. For this reason, the realization of the equipment is dependent. This is the most important problem of ANN. When ANN produces a probing solution, it does not give a clue as to why and how. This reduces trust in the network. Further drawbacks associated with ANNs include precipitation prediction and peak-value prediction.

[2] Training the network using Multilayer Perceptron (MLP)

The vast majority of ANN models for flood prediction are often trained with a back-propagation neural network. Simplicity, nonlinear activation, and a high number of layers are characteristics of the MLP. Due to these characteristics, the model was widely used in flood prediction and other complex hydrogeological models. In an assessment of ANN classes used in flood modeling, MLP models were reported to be more efficient with better generalization ability. Nevertheless, the MLP is generally found to be more difficult to optimize. Back-percolation learning algorithms are used to individually calculate the propagation error in hidden network nodes for a more advanced modeling approach. The output values of a perceptron can take on only one of two values (0 or 1) due to the hard-limit transfer function.

Perceptrons can only classify linearly separable sets of vectors. If the vectors are not linearly separable, learning will never reach a point where all vectors are classified properly.

[3] Predictive modeling using Decision Trees (DT)

The machine learning method of DT is one of the contributors in predictive modeling with a wide application in flood simulation. DT uses a tree of decisions from branches to the target values of leaves. In classification trees, the final variables in a DT contain a discrete set of values where leaves represent class labels and branches represent conjunctions of features labels. When the target variable in a DT has continuous values and an ensemble of trees is involved, it is called a regression tree. DTs are classified as fast algorithms; they became very popular in ensemble forms to model and predict floods. The classification and regression tree, which is a popular type of DT used in ML, was successfully applied to flood modeling. Further DT algorithms popular in flood prediction include reduced-error pruning trees, Naïve Bayes trees, chi-squared automatic interaction detectors, logistic model trees, alternating decision trees etc. Despite this there are many drawbacks to using DTs. A small change in the data can cause a large change in the structure of the decision tree causing instability. For a DT sometimes calculation can go far more complex compared to other algorithms. And it takes a lot of time to train a model.

[4] Extracting Information from Sources with Wavelet Neural Networks

Wavelet transform (WT) is a mathematical tool which can be used to extract information from various data sources by analyzing local variations in time series. Wavelet transforms supports the reliable decomposition of an original time series to improve data quality. The accuracy of prediction is improved through discrete WT (DWT), which decomposes the original data into bands, leading to an improvement of flood prediction lead times. DWT decomposes the initial data set into individual resolution levels for extracting better-quality data for model building. DWTs, due to their beneficial characteristics, are widely used in flood time-series prediction. In flood modeling, DWTs were widely applied in, e.g., rainfall–runoff, daily stream-flow, and reservoir inflow. Furthermore, hybrid models of DWTs, e.g., wavelet-based neural networks, which combine WT, feed-forward neural network, and wavelet-based regression models, which integrate WT and multiple linear regressions, were used in time-series predictions of floods.

[5] Adaptive Neuro - Fuzzy Inference System (ANFIS)

The fuzzy logic is a qualitative modeling scheme with a soft computing technique using natural language. Fuzzy logic is a simplified mathematical model, which works on incorporating expert knowledge into a fuzzy inference system (FIS). An FIS further mimics human learning through an approximation function with less complexity, which provides great potential for nonlinear modeling of extreme hydrological events, particularly floods. Adaptive neuro-FIS, or so-called ANFIS, is a more advanced form of neuro-fuzzy based on the T–S FIS, first coined. Today, ANFIS is known to be one of the most reliable estimators for complex systems. ANFIS technology, through combining ANN and fuzzy logic, provides higher capability for learning. This hybrid ML method corresponds to a set of advanced fuzzy rules suitable for modeling flood nonlinear functions. But it comes with some drawbacks. The computational cost of ANFIS is high due to complex structure and gradient learning. This is a significant bottleneck to applications with large inputs. Moreover, in terms of interpretability, ANFIS with grid partitioning produces a large number of rules which indeed cannot be easily understood by model users. Hence, interpretability is highly compromised, even though, the large number of rules contribute to improvement in model accuracy. Additionally, the trade-off between interpretability and accuracy is considered as crucial problem.

[6] Reducing Expected Errors with the help of Support Vector Machine

Hearst Et Al proposed and classified the support vector (SV) as a nonlinear search algorithm using statistical learning theory. Later, the Support Vector Machine (SVM) was introduced as a class of SV, used to minimize over-fitting and reduce the expected error of learning machines. SVM is greatly popular in flood modeling; it is a supervised learning machine which works based on the statistical learning theory and the structural risk minimization rule. The training algorithm of SVM builds models that assign new non-probabilistic binary linear classifiers, which minimize the empirical classification error and maximize the geometric margin via inverse problem solving. SVM is used to predict a quantity forward in time based on training from past data. A common disadvantage of non-parametric techniques such as SVMs is the lack of transparency of results. It is neither a linear combination of single financial ratios nor has it another simple functional form. The weights of the financial ratios are not constant.

[7] Ensemble Prediction Systems (EPSs)

A multitude of ML modeling options was introduced for flood modeling with a strong background. Thus, there is an emerging strategy to shift from a single model of prediction to an ensemble of models suitable for a specific application, cost, and dataset. ML ensembles consist of a finite set of alternative models, which typically allow more flexibility than the alternatives. Ensemble ML methods have a long tradition in flood prediction. In recent years, ensemble prediction systems (EPSs) were proposed as efficient prediction systems to provide an ensemble of N forecasts. In EPS, N is the number of independent realizations of a model probability distribution. EPS models generally use multiple ML algorithms to provide higher performance using an automated assessment and weighting system. The advantage of EPS is the timely and automated management and performance evaluation of the ensemble algorithms. Therefore, the performance of EPS, for flood modeling in particular, can be improved. The disadvantages of this approach are that it relies heavily on observed stream-flow data and requires regular updates with new data. Another drawback of the method is that the uncertainty is not captured equally across the water level spectrum.

CHAPTER 3

SOFTWARE REQUIREMENT ANALYSIS

Software Requirement Analysis in the field of systems engineering and software engineering, encompasses those tasks that are used for determining the needs or conditions to meet for a new or altered product or project, taking account of the possibly conflicting requirements of the various stakeholders, analysing, documenting, validating and managing software or system requirements.

3.1 Feasibility Study

The main objective of the feasibility study is to treat the technical, operational and economic feasibility of developing the application. Feasibility is the determination of whether or not project is worth doing. The process followed in making this determination is called feasibility study. All systems are feasible, given unlimited resources and infinite time. The feasibility study to be conducted for this project involves:

- Technical Feasibility
- Operational Feasibility
- Economic Feasibility

3.1.1 Technical Feasibility

It is the measure of the specific technical solution and the availability of the technical resources and expertise. It is one of the first studies that must be conducted after a project has been identified. A technical study of feasibility is an assessment of the logistical aspects of business operation. This is considered with specifying equipment and software that will successful satisfy the user requirement. The technical needs of the system may vary considerably but should include the facility to produce outputs in a given time, response time under certain conditions and the ability to process a certain amount of transaction at a certain speed.

The proposed system is developed by using Juypter Notebook software. The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modelling, data visualization, machine learning, and much more. The "notebook" term can colloquially make reference to many different entities, mainly the Jupyter web application, Jupyter Python web server, or Jupyter document format depending on context. A Jupyter Notebook document is a JSON document, following a versioned schema, containing an ordered list of input/output cells which can contain code, text, mathematics, plots and rich media.

3.1.2 Operational Feasibility

Operational feasibility is mainly concerned with issues like whether the system will be used if it is developed and implemented, whether there will be resistance from the users which will affect the possible application benefits. It is the ability to utilize, support and perform the necessary tasks of a system or program. It includes everyone who creates, operates or uses the system. It is the measure of how well a proposed system solves the problem and takes advantages of the opportunities identified during the scope definition and problem analysis phases. This system helps in many ways. It reduces the burden of maintaining bulk of records of all the rainfall data. Maintenance of the project is also easy and understandable and no major training and new skills are required.

3.1.3 Economic Feasibility

Economic feasibility is the most frequently used method for evaluating the effectiveness of the new system. Economic feasibility is the measure of the cost effectiveness of an information system solution. Without a doubt, this measure is most often and important one of the three. Information systems are often viewed as capital investments for the business, and, as such, should be subjected to the same type of investment analyses as other capital investments.

Economic analysis is used for evaluating the effectiveness of the proposed system. In economic feasibility, the most important is cost-benefit analysis. This project is economical as it mainly depends on the software components which are freely available.

CHAPTER 4

SYSTEM REQUIREMENT SPECIFICATION

The main purpose of System Requirement Specification is to translate the ideas in the minds of a client into a formal document. Through System Requirement Specification the client clearly describes what it expects from the proposed system and the developer clearly understands what capabilities are required to build the system. It includes a variety of elements (see below) that attempts to define the intended functionality required by the customer to satisfy their different users. The purpose of this document is to serve as a guide to developers and testers who are responsible for the development of the system.

4.1 Functional Overview

- Rainfall data is collected from meteorological websites.
- An algorithm is used for creating a model.
- The chances of a flood are predicted.

4.2 Operating Environment

Operating environment involves minimum software and hardware requirements required by the system.

4.2.1 Software Requirements

- Operating System : Windows 7 or above.
- Tools used : Jupyter 6.0 or above.
- Programming Language : Python Programming Language.

4.2.2 Hardware Requirements

- Processor : Pentium 4 or above
- RAM : 6GB or above
- Hard Disk : 10GB or above
- Input device : Standard Keyboard and Mouse.
- Output device : High Resolution Monitor

4.3 Functional Requirements

Functional Requirements defines the function of a system or its component. A function is described as a set of inputs, the behavior and outputs. Functional requirements specify particular results of a system. Functional requirements drive the application architecture of a system. Following are the functional requirements used in the project.

- Process data in any required format i.e., xls,csv etc.
- Import the data and store its metadata relating to the segmentation without corrupting the contents.
- Concepts and target extraction.
- Predicting the outcome.

4.4 Non-Functional Requirements

Non-Functional requirement is a requirement that specifies criteria that can be used to judge the operation of a system, rather than specific behaviors. Non-Functional requirements are often called as quality attributes of a system. The following are the nonfunctional requirements of the application.

- Availability: The System will work as required according to the specified requirement.
- Reliability: The System has the ability to consistently perform the intended or required functions.
- Maintainability: The maintenance of a functional unit can be performed in accordance with the prescribed requirements
- Accessibility: The System can be accessed by any appropriate users.
- Operability: The System has the ability to keep itself safe and reliable, according to predefined operational requirements.
- Usability: The System is ease of use and learnable by the users.
- Responsiveness: The system has ability to respond to the user very fast as soon as the input is fed.
- Cost: The system is economically feasible.

4.5 Performance Requirements

This application system will avoid use of papers and going through each and every detail for each state or district. Thus, this application results in efficient usage of time. This application will be useful for individuals, city planners, researchers etc. Application calculates the result automatically and displays it to the user. Interface is designed in such a way that user can understand very easily. Maintenance of the project is easy and understandable.