Studies in Big Data 118

Aboul Ella Hassanien
Ashraf Darwish *Editors*

The Power of Data: Driving Climate Change with Data Science and Artificial Intelligence Innovations

Springer

Studies in Big Data

Volume 118

Series Editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland

The series "Studies in Big Data" (SBD) publishes new developments and advances in the various areas of Big Data- quickly and with a high quality. The intent is to cover the theory, research, development, and applications of Big Data, as embedded in the fields of engineering, computer science, physics, economics and life sciences. The books of the series refer to the analysis and understanding of large, complex, and/or distributed data sets generated from recent digital sources coming from sensors or other physical instruments as well as simulations, crowd sourcing, social networks or other internet transactions, such as emails or video click streams and other. The series contains monographs, lecture notes and edited volumes in Big Data spanning the areas of computational intelligence including neural networks, evolutionary computation, soft computing, fuzzy systems, as well as artificial intelligence, data mining, modern statistics and Operations research, as well as self-organizing systems. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution, which enable both wide and rapid dissemination of research output.

The books of this series are reviewed in a single blind peer review process.

Indexed by SCOPUS, EI Compendex, SCIMAGO and zbMATH.

All books published in the series are submitted for consideration in Web of Science.

Aboul Ella Hassanien · Ashraf Darwish
Editors

The Power of Data: Driving Climate Change with Data Science and Artificial Intelligence Innovations

Springer

Editors
Aboul Ella Hassanien
Faculty of Computers and Artificial Intelligence
Cairo University
Cairo, Egypt

Ashraf Darwish
Faculty of Science
Helwan University
Cairo, Egypt

ISSN 2197-6503 ISSN 2197-6511 (electronic)
Studies in Big Data
ISBN 978-3-031-22455-3 ISBN 978-3-031-22456-0 (eBook)
https://doi.org/10.1007/978-3-031-22456-0

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The UN COP27 2022 conference in Egypt on climate change offers opportunities for leaders of the world to take action and make meaningful and urgent commitments and strategies to reduce greenhouse gas emissions to control global temperatures. The climate change impact is already threatening human existence across the universe. Climate change is one of the humanity's major challenges in the current century. It will continue to be a major problem in the next century not only due to its long-term effects on sustainable development, its pervasive nature, complexity, and consequences such as environmental hazards such as drought, desertification, and healthcare and agricultural problems. Climate change is used to describe the alteration in the climate over time caused by human activities and some naturally stirring events such as earthquakes and explosions. As a result, global warming is caused by the variabilities in climate variables which are occasioned mostly by human activities. Several current research studies have predicted that if immediate action is not made to address the impacts of the changes in the climate, the implications could lead to a lack of food and water in the next decades, as well as the effect on the health of people.

As a result, the current situation necessitates a quick response from all parties involved in devising, adopting, and implementing the essential steps and adaptation for reducing these potential hazards through data science and emerging and innovative technology.

Finding sustainable solutions to this issue will significantly affect the global, regional, national, and local community levels. The urgency in finding technological solutions to this global problem will require effective actions from governments, the industrial level, the private sector, and civil society.

There is an increasing recognition that data science and emerging and innovative technologies will play a major role in the national and global strategies to combat climate change challenges. Therefore, these technologies can potentially present solutions to complex development challenges brought by climate change. Data science and emerging digital and innovative technologies such as artificial intelligence, the Internet of things, digital twins, blockchain, drones, cloud computing, and sensor networks have drawn significant attention in the last years to tackle climate change challenges and reduce greenhouse gases and their impact. The impact and deploying these technologies on human life have witnessed a great interest in some areas related to climate change adaptation and mitigation, such as renewable energies, ocean dynamics, environmental quality control, biodiversity, agriculture, water sustainability, and energy sustainability. In addition, emerging digital technologies can deliver more efficient, rapid, and reliable decision-making solutions before, during, and after the occurrence of hazards as a consequence of climate change. These technologies can provide a chance to deliver sustained solutions to many societal challenges relating to climate change. Therefore, data science and emerging and innovative digital technologies can improve resilience to global warming-related, natural hazards, reduce emissions, and enhance humans' ability to take the necessary steps to realize net zero.

The contents of this book are divided into four parts.

Part "Artificial Intelligence in Climate Change Applications" is devoted to the applications of artificial intelligence for floods crises prediction based on machine learning techniques, the impact of climate change on air flight emissions for green aviation, the prediction of water quality index using machine learning techniques, the impact of artificial intelligence on waste management for climate change, and finally, a machine learning-based model for predicting temperature under the effects of climate change.

Part "Emerging Technologies in Industry and Energy Sector" concerns are emerging technologies' role in the climate change industry and energy sector. It includes the prediction of emissions of cars based on artificial intelligence techniques, a decision support framework for photovoltaic renewable energy prediction, a clean energy management system based on the Internet of things and sensor networks, digital twins for the energy management sector, using a case study based on Internet of things to design ozone prediction model for mitigating the effect of climate change, and finally suing emerging technologies for sustainable zero energy in buildings using multistage optimization in Egypt.

Part "Emerging Climate Change Technology in Agriculture Sector" is devoted to presenting the applications of emerging technologies in the agriculture industry. It concludes the crop prediction with climate factors and soil properties based on an intelligent recommendation mode, an intelligent recommendation model for strategic crops in Egypt using deep learning, a data-driven decision support system for an innovative water system for climate change problems, and finally, the role of artificial intelligence in water management systems for irrigation in agriculture.

Part "Emerging Climate Change Technologies in Healthcare Sector" presents the application of emerging technologies in the healthcare sector to tackle climate change problems. It includes the influence of climate change on the re-emergence of malaria using artificial intelligence.

The authors of this book wish to acknowledge the encouragement of the organizing committee of COP27, who motivated them to prepare this book. The success of publishing this book is first of all a result of the quality and the motivation of its

authors and researchers. So, we would like to thank all participants of COP27 and researchers and authors of chapters of this book for their contributions. We are also very grateful to the reviewers, whose very consistent reviewing of manuscripts was of great help in improving the quality of many chapters of this book. We also owe our gratitude to the publisher of this book for their willingness to deal with the chapters of this book.

Cairo, Egypt

Prof. Aboul Ella Hassanien
Prof. Ashraf Darwish

Contents

Artificial Intelligence in Climate Change Applications

Artificial Intelligence for Predicting Floods: A Climatic Change Phenomenon

Mohamed Torky, Ibrahim Gad, Ashraf Darwish, and Aboul Ella Hassanien

1 Introduction

Climate change is fast becoming one of the biggest and most significant challenges at this time. Many countries in the world are now witnessing radical changes in the features of their climate features. For instance, rainfall styles are shifting, temperatures are increasing, wildfires are starting to occur more frequently, snow and glaciers are melting, the global mean sea level is rising, and floods are becoming more and more dangerous to human life [1]. Climate changes we face confirm that the recent strategies of the world to manipulate them are not up to the challenge we face. For example, these policies will lead to a global temperature rise of + 3 °C. This puts our planet at real risk from more severe climate change effects before the end of the century [2]. Climate change is the large-scale, long-term shift in weather patterns, in particular, due to global warming phenomena. For instance, to make Earth's temperature below the + 2 °C target, CO_2 emissions must be decreased and be close to zero by 2070, with total gas emissions turning negative thereafter [2].

M. Torky (✉)
Faculty of Artificial Intelligence, Egyptian Russian University, Badr City, Egypt
e-mail: mtorky86@gmail.com

I. Gad
Faculty of Science, Tanta University, Tanta, Egypt
e-mail: ibrahim.gad@science.tanta.edu.eg

A. Darwish
Faculty of Science, Helwan University, Cairo, Egypt
e-mail: ashraf.darwish.eg@ieee.org

A. E. Hassanien
Faculty of Computers and AI, Cairo University, Giza, Egypt
e-mail: aboitcairo@cu.edu.eg

M. Torky · I. Gad · A. Darwish · A. E. Hassanien
Scientific Research Group in Egypt (SRGE), Giza, Egypt

A. E. Hassanien and A. Darwish (eds.), *The Power of Data: Driving Climate Change with Data Science and Artificial Intelligence Innovations*, Studies in Big Data 118,
https://doi.org/10.1007/978-3-031-22456-0_1

On the other hand, flooding is another climatic phenomenon that is possible to occur due to the change in water cycles on our planet. So, the important question is: How does climate change affect floods? Since climate changes occur due to the burning of oil, coal, and gas, this leads to that the weather will change to become warmer, and wetter. You can simulate this scenario by putting some water in a pan and heating it. Increasing temperatures make the water boils, and the steam raises and makes it more moisture, this leads to the constitution of dense clouds that cause heavy rainfalls and occur flooding. There are five climatic causes of flooding: heavy rains, storms, overflowing rivers, lack of vegetation, and melting snow and ice [3]. A more moisture atmosphere can lead to more intense downpours in short times; this can raise the risk of sudden and un-expecting flooding. Moreover, increasing the atmosphere temperature means there is more heat for climate systems that cause heavy rainfalls. Floods shaped the greatest proportion of economic destruction from the unstable climate in Australia over the last 10 years, followed by droughts, and tropical. If these levels of carbon emissions continue without drastic solutions, the Australian economy will lose $40 billion per year by 2060 due to floods [4]. Floods not only have a serious impact on the global economy, and people's life and, but also on our beloved animals, valuable livestock, and wildlife [5]. Hurricane strength is also forecasted to rise as the weather continues to warm. Intense storms can lead to a greater risk of coastal flooding from storm attacks, a danger that will be further growth by sea level rise. Hurricanes occurrences have also serious impacts on moving floods to drink water systems and wastewater services, which can damage treatment procedures or water distribution systems and all infrastructure lines [6].

Flood events are also related to changing rain patterns on snow. In some cases, the downpour-on-snow events occur while the soil is still partially congealed. The frozen soil, which is already saturated, cannot absorb additional water. Therefore, the more rainfall the more snowmelt runoff. This leads to creating many flood rivers with a huge amount of flowing water. This scenario of downpour, frozen soils, and snowmelt was the major reason for Midwest flooding in March 2019 that caused over US$12 billion in damage. In addition, In June 2022, a huge flooding event hit the western mountains series in the Western U.S, where snowmelt combined storms and rainfall dumped up to five inches of rain over three days in Yellowstone National Park and surrounding places, rapidly melting snowpack. Huge amounts of flowing water shaped a dystrophic flood that damaged utilities, and roads, and forced more than 10,000 people to evacuate [7].

Warmer air and wildfires are other reasons for global warming and heavy rains, which generate complex changes in our environmental ecosystems and increase the potential of flood events in various geographic areas on our planet.

Green technology (or environmental technology/green tech.) may be a radical solution to many climate change challenges [8]. Green tech. refers to using science and technology to mitigate the negative environmental impact of human activity and protect the world's natural resources. Although this technology is relatively young, it has fast become a great interest among scientists, engineers, and politicians due to increasing awareness about the impacts of climate change and the exhaustion of environmental resources. The green tech involves four solutions to mitigate climate change and minimize the global warming that increases the potential of flood events [9]:

1. *Low-carbon emission*: It is estimated that our classical homes and building account for 38% of global warming gas emissions around the world [10]. On the other hand, low-carbon houses are designed to require minimal cooling and heating and produce very little waste and pollution. This certainly minimizes the global warming that causes heavy rains and floods.
2. *Global warming reversing*: This process is based on the idea of the possibility of storing and capturing carbon by pulling it from the atmosphere and using it to make synthetic fuel. This solution will help in minimizing atmosphere temperature and global warming. Therefore, the chances of precipitation causing floods are diminishing. Although global warming reversing is still a more expensive task, some studies expect that the costs of direct air carbon storage could come down by a multiple of six, making it a much more achievable process [11].
3. *Renewable energy storage*: storing clean energy such as solar energy for a long time will minimize the dependence on fossil fuels. Burning fossil fuels launch large amounts of CO_2, a greenhouse gas, into the air. Greenhouse gases trap heat in our atmosphere, causing global warming which is a major motivation for rainfall and occurring flooding.
4. *Green Hydrogen Production*: Battery-powered electric vehicles are a common type of electric car, but there's another type of electrical vehicle, which is called a *fuel cell electric vehicle.* These vehicles don't run on batteries, they run on green hydrogen. The main advantage of fuel cell electric vehicles (i.e. Hydrogen-based) is that they don't produce any harmful emissions, and work more efficiently than burning-powered vehicles, and. experts forecast that by 2050, green hydrogen will eventually run over 420 million vehicles and more than 20% of passenger ships [12]. Hydrogen fuel is still expensive, but hydrogen-powered vehicles are already used in many countries all over the world and could become crucial for cutting transport emissions and global warming from fossil fuels-based vehicles.

Artificial Intelligence (AI) is an important technology that can assist in using green technology for mitigating many climates change challenges. Experts and research community and experts have already started focusing on climate informatics with AI models. AI models have a greater potential to predict, assess, classify, recognize, and mitigate climate change hazards with the efficient use of sensing devices, big environmental data patterns, and learning algorithms [13–15]. AI has many diverse applications in climate change areas. For example, AI-based prediction models and decision support systems (DSS) for weather prediction, AI-based machine vision applications in climate informatics, AI-based assessment models to promote eco-friendly energy production and consumption, AI-based prediction models for earth hazard management, AI-based assessment models for reducing the impacts of global warming, Deep learning models for earth informatics and sustainable earth surveillance, AI-based assessment models to minimize carbon footprints for a sustainable environment [16].

Flood prediction is another application of artificial intelligence in climate change area [17]. In flood prediction, predicting flood hazard variables using physical methods can involve a chain of hydraulic and hydrologic models that describe the physical features of floods. Although such models provide flood understanding, they often require high computational and data details. Therefore, Machine-learning techniques can be used as an alternative solution to flood prediction issues. They have the potential to enhance accuracy as well as decrease calculating time and model development cost. This chapter is designed to discuss how AI can be utilized for developing machine/deep learning models to predict flood events. Moreover, the chapter introduces a proposed AI model to add a novel solution to this issue.

The remaining sections of this chapter can be organized as follows. Section 2 presents background about artificial intelligence and flood prediction issue. Section 3 discusses the proposed AI model for predicting flood events. Section 4 presents the experimental results. Section 5 concludes this work.

2 Artificial Intelligence Approaches and Flood Prediction

The recent classical approaches used for flood prediction do not provide accurate forecasting results because of the shortage of sufficient data. Usually, the inaccurate prediction results of those physical approaches lead to catastrophic results as the people cannot timely make the optimal evacuation decisions and put their lives in big danger. One of these traditional approaches is Hydrology [18]. Hydrology is a physical forecasting method that studies the linear relationship between rainfall falling-depth in each time and the resulting water runoff. The sum of this incremental function is then used to estimate the flood flow hydrograph. Another physical method to forecast floods is based on Hydraulic principles [19]. In this way, flood prediction is estimated by studying the relationship between the rainfall pressure on water surfaces and the generated equal forces that push water in all directions.

Figure 1 illustrates how meteorological and hydrological systems are working together for forecasting floods. Note that, all figure's links and rectangles may be required or used. However, the fuzzy data of hydro-meteorological variables such as downpour, water level, flow, and rainfall also result in fuzzy and inaccurate prediction results.

Artificial Intelligence models can do marvels when utilized to anticipate future events such as floods. AI models can mimic non-linear behaviors of a given phenomenon and produce more precise- forecasting results. AI algorithms can predict flood events before they occur, thanks to huge volumes of high-quality datasets. AI-based flood prediction models have some strengths over classical hydrological and hydraulic modeling: (1) Ai-models can provide prediction results and warnings in milliseconds. (2) They work on real observed data and did not consider any hypothetical observations or data structures. (3) They are self-learning methods that can

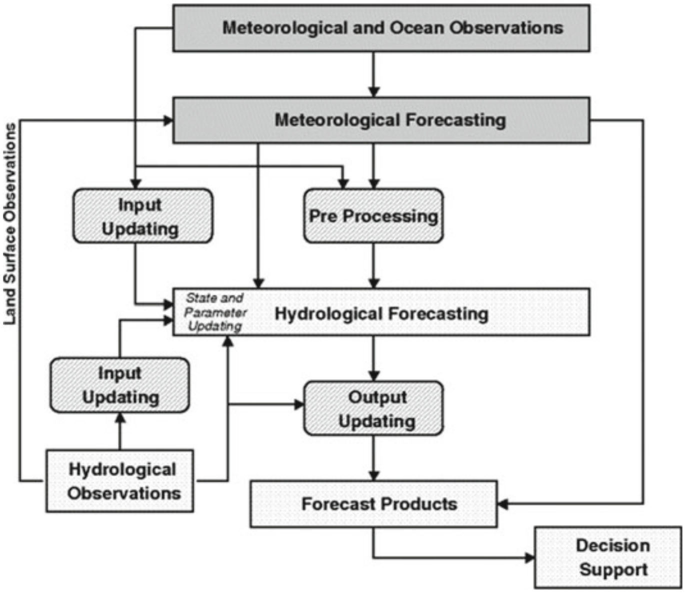

Fig. 1 Meteorological and hydrological flood prediction systems

improve prediction results with time and data. Googles experts believe that AI-flood prediction models can assist in producing flood warnings in a very short time compared to the physical methods and hence can avoid 30 to 50% damage [20].

2.1 How Does AI-Models Work to Forecast Floods

Data is the major requirement for applying AI models. Therefore, in flood prediction issues, sensor devices and satellites are excellent tools for gathering various climatic data patterns from various positions. Based on the collected data, AI algorithms can analyze vectors of data features (e.g. Temperature, rainfall, pressure, wind speed, direction, see and oceans' status, etc.) of this massive information and provide anticipation results about the time and place of flood events.

The methodology of AI-models starts with training (or learning) the used AI-model with the collected dataset (e.g. Climatic data, past rainfall data, temperature, pressure, wind, etc.). Using precise real-world data gathered from governmental and meteorological agencies, the AI- models provide the required flood prediction results. Figure 2 illustrates the AI methodology for flood prediction. The meteorological dataset is input to an AI system as *training data*, which is used to learn a given AI model how to predict flood events based on an extracted vector of data features. After the learning process is terminated, an AI model is then tested with different data patterns (i.e. *testing data*) to verify its accuracy in predicting flood events. The

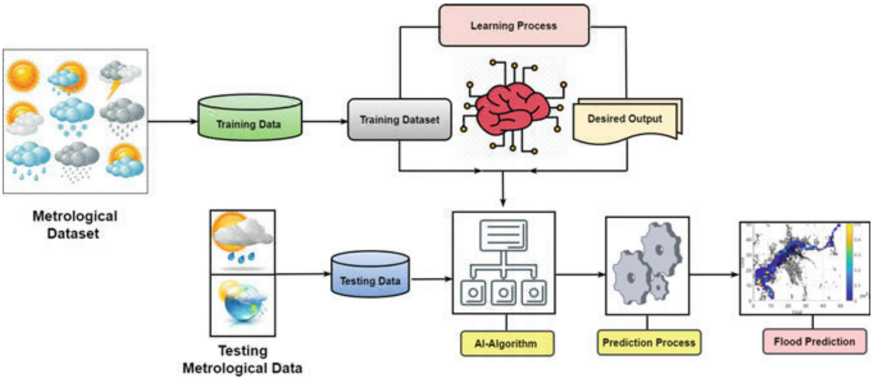

Fig. 2 AI methodology for predicting flood events

learning process in AI is the most significant step with an AI prediction model. The more data volumes the more the learning accuracy of a given AI model.

There are three common learning processes in AI: (1) *supervised learning* is machine learning that requires labeled datasets to train AI algorithms to classify data or predict outputs. (2) *unsupervised learning* is another kind of machine learning in which an AI model is provided with unlabeled and unstructured data. The main objective of this approach is to find structure/patterns in the input data; therefore this learning scheme is called self-learning as an AI algorithm can detect the hidden patterns in unlabeled datasets. (3) *Reinforcement learning* is another learning methodology that reacts to a dynamic environment in which it must achieve a certain goal without any labeled data that can help it to achieve this goal. Therefore, learning a model is based on transactions with a specific domain.

A novel AI-models that can be used for predicting flood events can be developed based on one of the previous learning methods or a hybrid of two of them. Regarding AI prediction models, by specifying only 2–3 features of a data pattern in the historical datasets (e.g. past flood events), the AI prediction model can forecast not only the time and position of flooding but also the seriousness of their occurrence.

Predicting flooding is hypothesized to achieve two goals, predicting flood water level, and predicting flood areas. Therefore, AI can employ two models for achieving these goals respectively, the hydrological Model, which is used to help scientists to predict water levels in rivers, and the invasion model, which determines locations most likely to be damaged by floods. This technique can assist in avoiding damage and mortalities. Moreover, gathered data patterns from social media such as Facebook, Twitter, and other smartphone apps can also help in tracking flooding. Social media-based data holds various data patterns, images, and videos that can help in determining flood locations by AI. Such schemes can monitor, and predict the damage caused by floods [21].

2.2 *Machine and Deep Learning for Flood Forecasting*

Artificial intelligence could help in developing warning and forecasting systems for forecasting flood events and resulting damage based on machine and deep learning algorithms [22]. Using machine-learning algorithms, floods can be predicted based on structured data, which is used to learn a machine-learning algorithm how to recognize key data features used to predict flood events. Machine learning algorithms can be categorized into three kinds of algorithms, supervised, unsupervised, and semi-supervised algorithms as clarified previously. Deep learning is a subset of machine learning that can perform more complex problems than machine learning based on artificial neural networks (ANN) designs. Therefore, Deep learning algorithms can predict floods based on big and unstructured data volumes in ways machine learning algorithms cannot easily do. Besides performing flood predictions, machine and deep learning algorithms can also monitor and map out/plot flood events while making decision support systems.

Designing machine and deep learning models for flood prediction specifically requires historical data of past flood events and real-time data of rainfall and other meteorological information. The sources of required real-time data involve multi-sensory devices such as radars and satellites, which can monitor the climate in a wide variety of locations. The best machine and deep learning models that can produce optimal results for flood prediction involve the following techniques:

(1) ***Artificial Neural Networks (ANNs)***: The neuron-layers design of ANNs makes them very dynamic, adaptable, and effective in learning to model complex flood predictions. These models can assist in flood forecasting floods with high fault tolerance and accurate approximation results. ANN algorithms can be fed with distributed rainfall data from various rain sensors fixed in multiple mountainous regions, then utilize this data to forecast the river water level in valleys between mountain regions. The flow of this river and its water level in these valley regions a few minutes before the flood occurs depict the river's behavior subject to sensor disruption [23]. This real-time flood prediction of an ANN model can be extremely helpful for timely warning and calamity prevention since it can give precise flood prediction within a few seconds. Therefore, ANNs are counted as one of the most dependable AI techniques for designing black-box models of complex and nonlinear relationships between rain flow and flood. As well as river flow and drainage prediction. Figure 3 depicts how ANNs design can be used for flood prediction. The learning process is powered with weights adjusted on neuron links until achieving the best accuracy of flood prediction results. The advance in using ANN models results in the development of deep neural network models able to enhance flood prediction results driven by radar observations and Image recognition based on integrating self-learning functions with neural network models. The highly reliable and precise flood prediction has been achieved by integrating hydrology and river engineering data to develop advanced deep learning models able to predict flood events accurately [24].

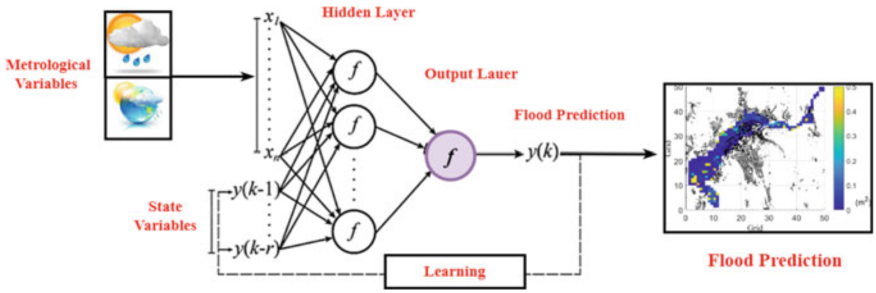

Fig. 3 Flood prediction using ANN model

(2) ***Neuro-Fuzzy Algorithms***: These types of algorithms are defined as mathematical models that employ neural networks and fuzzy logic rules in developing novel machine learning flood prediction models. In this approach, flood knowledge can be processed using Neuro-Fuzzy Inference System (ANFIS) [25]. The learning process of this approach is based on an approximation-mathematical function in developing a nonlinear model for predicting hydrological events like floods. ANFIS is a popular and efficient flood forecasting approach due to its fast programming, accurate learning, and powerful abilities for standardization. The ANFIS methodology involves five layers of calculations, input layer, fuzzification layer, inferences layer, defuzzification layer, and output layer as depicted in Fig. 4. The Input metrological variables are to be fuzzified in layer 2. The inference process and fuzzy rules are executed in layer 3. Variables defuzzification of each corresponding rule is carried out in layer 4, and finally, in layer 5, the final flood prediction results are output.

(3) ***Statistical Learning***

Statistical learning models are popular data-driven techniques, which have become increasingly used for event regression analysis. These types of models

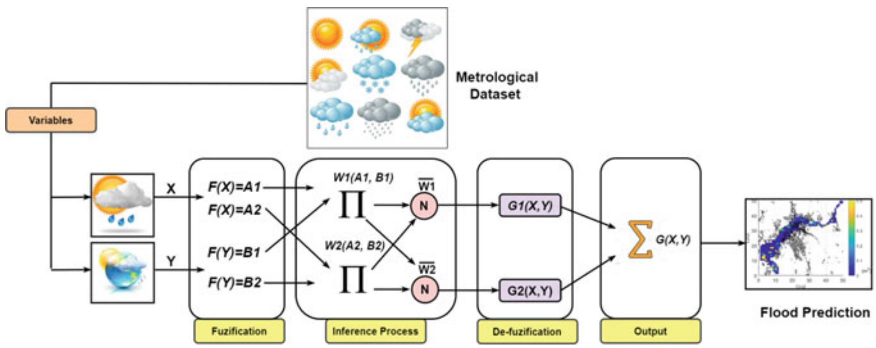

Fig. 4 Flood prediction using ANFIS approach

gave the ability mimic nonlinear hydrological functions. Support Vector Regression (SVR) [26] and Support vector machine (SVM) [27] models are powerful supervised learning models, and their methodology is based on statistical learning theory and the operational risk reduction rule. This makes these models effective techniques to predict flash flood events at different lead times and can serve as an alternative solution to ANNs. The statistical learning–based models can be utilized in various flood prediction cases such as precipitation, extreme rainfall, rainfall-runoff, stream flow, reservoir inflow, soil and weather moisture, and flood quantities and time series. Therefore, SVR and SVM both give promising results, better performance, and remarkably standardized methodologies as compared to ANN.

2.3 AI Systems for Flood Forecasting

Some large companies specializing in information technology and artificial intelligence have begun to develop complete systems and applications that can be greatly relied upon in predicting flood events, warning, and evaluating the risks resulting from the expected occurrence of the flood. These applications can also serve as guidance and recommendation systems for citizens and decision-makers to take the necessary measures to avoid the destruction caused by flood occurrences. In this subsection, we briefly highlight two projects developed as mature systems used in predicting flood events. One of these systems is developed by *Google*, and the other one is developed by *Piccard.AI* organization.

A. *Google's AI-powered Flood Prediction System*
Google, the giant of technology has launched an AI-based system in 2017 for predicting river flood events in south Asia [28]. China, India, and Bangladesh are among the regions that are greatly hit by floods and natural disasters with huge damage to infrastructure and loss of lives. Therefore, Google's AI experts developed reliable and scalable flood prediction models depending on machine learning technology. The main idea of Gogol's flood prediction system is based on identifying a river's probability to make flooding based on satellite images and river trajectory. Moreover, Google's system is scalable to involve a set of hydrological variables such as water level, and precipitation to enhance prediction accuracy and extend the lead-time of flood prediction. The pilot version of this system proved that it was able to predict river floods with an accuracy of 75%. Another important advantage of Google's flood prediction system is it can deliver precise real-time flood information and warns the citizens through Google search engine and Google Maps to take defensive precautions. After applying several pilot versions of this system in various regions, Google announced that its flood prediction system issued more than 100 million warning notifications, saving the lives of 360 million citizens in the areas most likely to be attacked by floods. Moreover, the system can launch flood alerts before flood emergence with a lead-time of 72 hours. Although the pilot versions of Google's flood system

proved satisfied with prediction results, it is limited to only river flooding. Therefore, Gogol's AI experts are recently working to update this system to be able to treat not only river floods but also "flash floods", and not only used in the regions covered by Internet area but also in poor regions that are not covered with internet services.

B. *Piccard's AI-based Flash Flood Warning and Prediction System*
Piccard's AI is a software association, which develops AI software systems for prediction, optimization, and automation problems. It has developed a flash flood prediction system able to help residents to avoid any socioeconomic damages. The main idea of this system is based on forecasting rainfall by using IoT sensors, and then, creating alerts before even the downpour occurs. The Piccard's AI system consists of four subsystems: (1) Water level surveillance devices fixed at various hotspot locations. (2) Set of machine learning techniques that can learn from climate data to forecast floods at those hotspot locations before flood occurrence within 1–4 hours. (3) A cloud-based mobile/desktop dashboard for analyzing and presenting flood behavior and its risks. (4) Email/SMS warning system of predicted floods to alert citizens in all the surrounding areas by the predicted flood [29]. The main advantage of this system is its ability to ideally analyze flood behaviors that cannot be addressed with drainage upgrades. By alerting citizens who are at risk, proactively locking roads and, authorities can minimize their damage and reduce the potential risks of flooding.

3 Proposed Machine Learning-Based Flood Prediction Model

In this section, a flood prediction model is proposed, which is based on utilizing machine learning and optimization algorithms. The proposed model consists of three phases: (1) flood data Pre-processing phase, and (2) feature selection based on the Pigeon swarm optimization (PSO) algorithm [30]. (3) Flood prediction based on machine learning phase. Figure 5 illustrates the three phases of the proposed model.

3.1 The Pre-processing Phase

One of the most known stages while developing machine-learning algorithms for data classification is cleaning the used data from any duplication, incompleteness, redundancy, etc. Without cleaning the train data, the classification models will produce accurate and biased results. One of the most popular automated methods that can do the data cleaning phase is SMOTE, which stands for "Synthetic Minority Oversampling Technique". This technique can automatically generate newly cleaned data using the old data. Table 1 shows a sample of the used flood dataset collected from

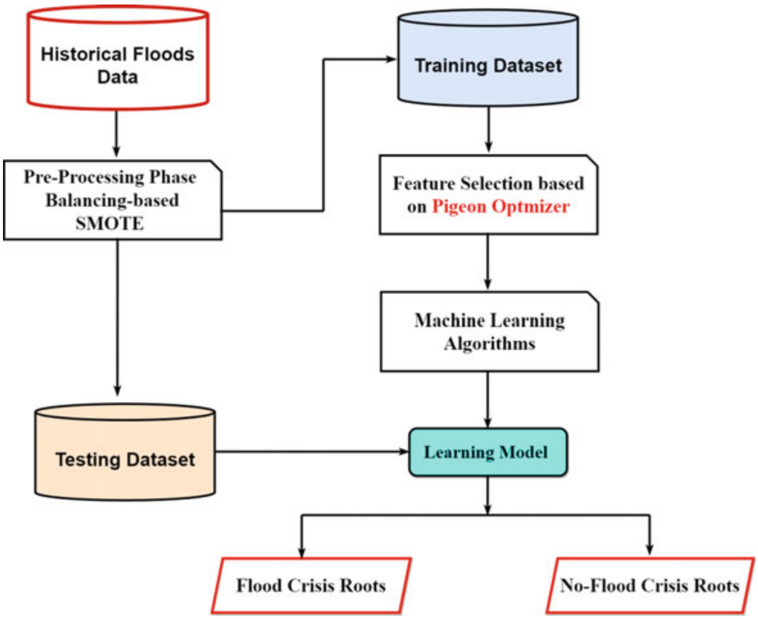

Fig. 5 Proposed machine learning-based flood prediction model

Kerala, India that describes the past floods from 1900 to 2020 [31]. The dataset consists of 120 samples of floods and 14 features. Table 2 shows counts of data labels before and after applying oversampling using SMOTE approach. Figures 1 and 2 shows the scatter plot that visualizes dataset classes before and after applying the SMOTE approach (Figs. 6 and 7).

3.2 Feature Selection Based on the PIO Phase

Pigeon-Inspired Optimization (PIO) is a novel swarm intelligence algorithm inspired by the homing behavior of pigeons [30]. The optimization methodology of the PIO algorithm can be utilized for finding the best solution to a given problem. In flood prediction problems, PIO can be used for finding the optimal reduct of features of the dataset that contribute to obtaining the most precise prediction results. This goal can be achieved by a fitness function to measure the validity of solutions and select the best features of an optimal solution. The used fitness function depends on the total number of features (TF), the performance of the classification model (accuracy), the reduct length, and three weight constants that contribute to the significance of prediction accuray, w_1, subscript and w_2, and w_3, where, $w_1 + w_2 + w_3 = 1$. Equation (1) defines the PIO-fitness function, while Eq. (2) defines the reduct weight that depends on the importance of the feature within the reduct. In our problem, we hypothesized

Table 1 Dataset sample

Subdivision	Year	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL RAINFALL	Floods
Kerala	1901	28.7	44.7	51.6	160	174.7	824.6	743	357.5	197.7	266.9	350.8	48.4	3248.6	Yes
Kerala	1902	6.7	2.6	57.3	83.9	134.5	390.9	1205	315.8	491.6	358.4	158.3	121.5	3326.6	Yes
Kerala	1903	3.2	18.6	3.1	83.6	249.7	558.6	1022.5	420.2	341.8	354.1	157	59	3271.2	Yes
Kerala	1904	23.7	3	32.2	71.5	235.7	1098.2	725.5	351.8	222.7	328.1	33.9	3.3	3129.7	Yes
Kerala	1905	1.2	22.3	9.4	105.9	263.3	850.2	520.5	293.6	217.2	383.5	74.4	0.2	2741.6	No
Kerala	1906	26.7	7.4	9.9	59.4	160.8	414.9	954.2	442.8	131.2	251.7	163.1	86	2708	No

Table 2 The count of data labels before and after applying SMOTE technique

Class label	Before SMOTE	After SMOTE (Oversampling)
Flood crisis	60	60
No-flood crisis	58	60

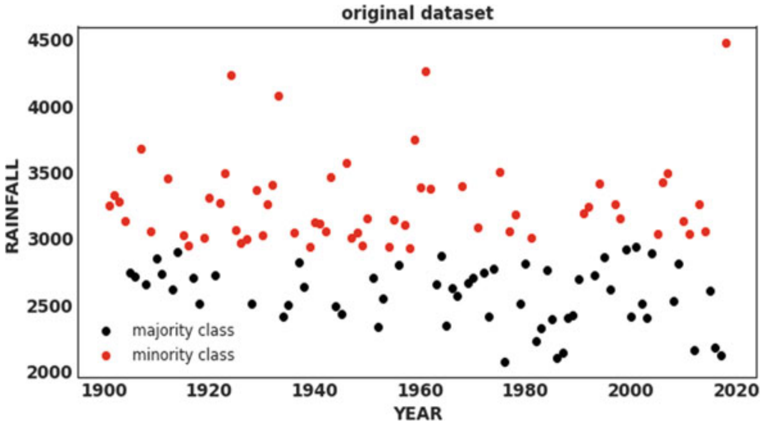

Fig. 6 The distribution of dataset classes **before** applying the SMOTE approach

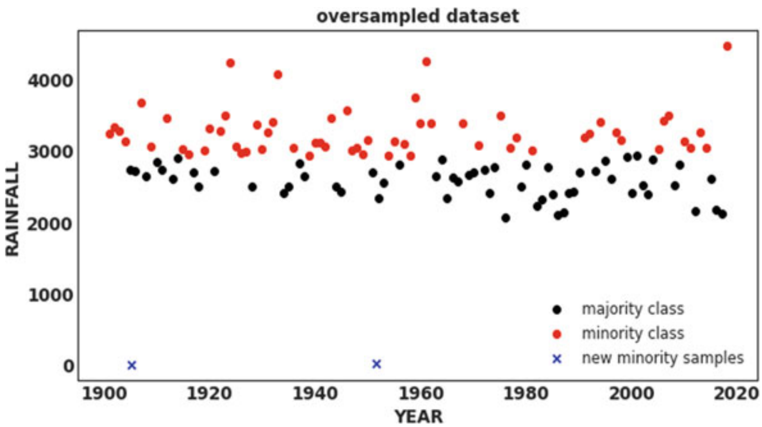

Fig. 7 The distribution of dataset classes **after** applying the SMOTE approach

that $w_1 = 0.70$, $w_2 = 0.22$, and $w_3 = 0.08$.

$$Fitness\ function = w_1 * accuracy + w_2 * F_1 + w_3 * \left(1 - \frac{SF}{TF}\right) \quad (1)$$

$$\text{Reduct weight} = \sum_{i} \text{importance}(\text{fi}_i) * \frac{\text{TF} - \text{SF}}{\text{TF}}, \text{such that importance}(\text{fi}_i) > \text{threshold} \tag{2}$$

3.3 Flood Prediction Based on Machine Learning Phase

In this phase, two learning models have been applied, called Gradient Boosting (GB) [32], and Random Forest (RF) [33], but with two different scenarios. In the first scenario, the GB algorithm is combined with the PIO algorithm for predicting floods, while, in the second scenario, the process of flood prediction is completely solved by the RF algorithm without using the PIO algorithm. The two scenarios have been applied and compared to identify the features of the flood crisis and predict its occurrence. The learning process is performed before and after applying SMOTE technique in the trained dataset to validate the efficiency of the proposed model. The performance is validated using the known performance metrics such as precision, recall, accuracy, and F1-score as in Eqs. (3, 4, 5, and 6) respectively.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{5}$$

$$\text{F1 - Score} = \frac{2(\text{Precision*Recall})}{\text{Precision} + \text{Recall}} \tag{6}$$

where, TP, TN, FN, and FP are truly positive, true Negative, False Negative, and False Positive numbers respectively.

4 Experimental Results

In this section, the effectiveness of the proposed flood prediction model is evaluated using a Kerala, India dataset. The proposed model has been developed in Python and the *scikit-learn* software packages. The testing method is carried out on a *Google Colab cloud* equipped with a CPU running at 2.6 GHz and 32 GB of RAM. Two scenarios of experiments have been carried out, and the obtained results can be detailed as follows.