Article

Not peer-reviewed version

Machine Learning for Flood Resiliency—Current Status and Unexplored Directions

Venkatesh Uddameri * and E. A. Hernandez

Posted Date: 22 April 2025

doi: 10.20944/preprints202504.1887.v1

Keywords: Flood Resiliency; Artificial Intelligence; deep learning; XAI; flood forecasting; flood life cycle; Deep learning; Recurrent Learning

 doi:10.20944/preprints202504.1887.v1

Article

Machine Learning for Flood Resiliency—Current Status and Unexplored Directions

Venkatesh Uddameri *,† and E. A. Hernandez †

Department of Civil and Environmntal Engineering, Lamar University, Beaumont, TX 77701
* Correspondence: vuddameril@lamar.ed; Tel.: +1-409-880-7207
† These authors contributed equally to this work.

Abstract: A systems-oriented review of machine learning (ML) over the entire flood management spectrum encompassing, fluvial flood control, pluvial flood management and resiliency-risk characterization was undertaken. Deep learners like long short-term memory (LSTM) networks perform well in predicting reservoir inflows and outflows. Convolution neural networks (CNN) and other object identification algorithms are being explored in assessing levee and flood wall failures. The use of ML methods in pump station operations are limited due to lack of public-domain datasets. Reinforcement learning (RL) has shown promise in controlling low impact development (LID) systems for pluvial flood management. Resiliency is defined in terms of vulnerability of a community to floods. Multi-criteria decision making (MCDM) and unsupervised ML methods are used to capture vulnerability. Supervised learning is used to model flooding hazards. Conventional approaches perform better than deep learners and ensemble methods for modeling flood hazards due to paucity of data and large inter-model predictive variability. Advances in satellite-based, drone-facilitated data collection and Internet-of-Things (IoT) based low-cost sensors offer new research avenues to explore. Transfer learning at ungaged basins holds promise but is largely unexplored . Explainable Artificial Intelligence (XAI) is seeing increased use and helps the transition of ML models from black-box forecasters to knowledge-enhancing predictors.

Keywords: flood resiliency; artificial intelligence; deep learning, XAI; flood forecasting; flood life cycle; deep learning; recurrent Learning

---

1. Introduction

Floods are said to occur when flows occur on an otherwise dry land or when the flow in a water body is well above normal values. There are many types of flood defined as per when and where they occur. Fluvial floods are known to occur when water levels in the channel flow over the banks and spill into the floodplain. Pluvial floods are caused by water directly accumulating on land surface that are not necessarily near a river. Coastal floods are caused when seawater is pushed onto land by high tides, storm surges and tsunamis. Urban floods are a type of pluvial flooding where flooding occurs in urban areas mainly due to excessive amount of impervious surfaces and insufficient drainage systems. Flash floods occur when a large amount of rain falls over a short-period of time. The high intensity of rainfall reduces the amount of infiltration and quickly leads to runoff. While not very common, the rise of groundwater levels in low-lying areas can also lead to flooding and is referred to as groundwater floods [1].

Flooding is a common natural disaster that affects every part of the Earth and causes billions of dollars of damage and even loss of life in many parts of the world. The US Department of Homeland Security states that 90% of the natural disasters in the US involve flooding [2]. Climate change models indicate that floods will become more intense and also increase in frequency in many parts of the world [3,4]. Increased urbanization to accommodate growing global population and heterogeneous economic growth is also diminishing natural infiltration capacity and increasing runoff and flooding in many areas [5]. In addition, improper construction of flood mitigation structures, dam breaks and

 doi:10.20944/preprints202504.1887.v1

other engineering failures can also lead to flooding [6]. Given the large-scale economic, environmental damages and health and loss of life hazards to humans, the deleterious impacts of floods must be prevented when possible or at least minimized to reduce damages.

It is important to ensure that flooding impacts on humans are either fully eliminated or controlled to the extent possible. Developing such strategies to combat flooding is not easy as the main drivers of flooding - 1) Weather and climate and 2) land use alterations cannot be predicted with a high degree of certainty and are prone to change. The impacts of floods must be managed so that the community can rebound back to normalcy as soon as possible. As societies continue to learn from flooding diasters - How have flood management paradigms changed over time? and what are the current approaches with regards to managing flooding risks?

Flood management requires decision-makers to have reliable data on various aspects. These data include, flooding characteristics such as peak flows, flood duration, flood volume, flood inundation depths. Factors causing flooding such as precipitation and land use alterations must be known as well. In addition, information on population distribution, social chracteristics, economic assets that are directly in the line of flooding hazards must also be mapped. Such data are not only useful for immediate flooding emergency response operations, but also to guide the development of new flooding infrastructure to absorb the shocks of flooding or at least mitigate their impacts to acceptable tolerances. In case of riverine flooding, the impacts on aquatic and riparian habitats and ecosystems are also of concern. Much of these data are hard to obtain and are not deterministic in nature. Therefore, decisions regarding flooding have to be made under considerable uncertainty.

Mathematical models have been used to support flood decision making. Models can be used to translate rainfall into runoff (for a given set of land use) and estimate the dimensions of flood such as peak flows in rivers, the area likely to be inundated, the depth of water on urban transportation infrastructure. These models are used not only for quantifying flooding related hazards, but also used to design flood control and flood mitigation structures, assess risks to humans and the environment due to flooding and develop new policies and guidelines that help in proactively combating the effects of flooding.

A wide range of tools and techniques have been proposed using a variety of different mathematical strategies. Mathematical models built using conservation principles of physics and at varying levels of spatial and temporal complexity have been available for some time now and continue to be widely used [7]. In a similar manner, a broad range of statistical tools have been used for flood estimation and forecasting [8,9], multivariate flood risk assessment frameworks [10,11]. In recent times, the use of machine learning (ML) methods is gaining ground [12]. ML methods are universal approximators that can capture highly nonlinear relationships. As such, their utility for flood forecasting has garnered much interest in recent times. The literature identified using a simple search "Flood and Machine Learning" has more than doubled from 19000 documents retrieved in 2015 to over 40,000 documents in 2024 by Google Scholar [13]. This change is indicative of the growing interest in using ML for flood forecasting and management studies.

The proliferation of machine learning methods in flood prediction applications has also led to several review papers in recent years [14,15]. However the emphasis of these reviews have largely been on prediction aspects. While predicting floods is indeed an important task, flood management also entails many other tasks such as ranking and prioritization of sites and susceptibility mapping. A review of machine learning usage in the context of flood management practices has not been undertaken to the best of the authors' knowledge. The main goal of this paper is to fill this important gap that currently exists in the literature.

The lack of a suitable review of machine learning methods for flood management also suggests that there are likely several unexplored directions of research. Identifying these unexplored directions will stimulate new avenues of research and address those aspects where machine learning methods can offer benefits that have not yet been brought to fruition. In a similar vein, newer studies will also make machine learning models more congruent with the needs of flood risk management.

To achieve the above-mentioned goals, the rest of the paper is organized as follows - 1) The methodologies adopted in this review are presented; 2) A brief review of flood management paradigms is undertaken to help readers understand how flood management has shifted from a more command-control approach to a more holistic resiliency paradigm that not only seeks to mitigate flooding hazards but also help bounce back to the original (or a better) state upon the cessation of flooding. To properly evaluate the benefits of machine learning it is also important to assess the strengths and limitations of physics-based models which have been used in flood studies over a longer period of time. The main physics-based modeling tools available today are briefly reviewed along with their strengths and limitations next. This evaluation seeks to set the stage of evaluating machine learning models. For sake of completeness, machine learning models are briefly described and their role in flood management is explored in greater detail.

2. Methodology

While the primary focus was on machine learning applications over the entire gamut of flood management, presenting a brief evolution of risk management frameworks from command-control type approaches to a community-focused resiliency approach was deemed necessary to place the review in proper context. Physics-based models provide a theoretically rigorous approach to flood modeling and continue to be used today. Figure 1, depicts the resiliency paradigm adopted in this study. Resiliency here is viewed as an umbrella of tools, technologies and policies that contribute towards a common goal of withstanding flooding threats or improving the ability of communities to quickly bounce back to normalcy or better after the flood.

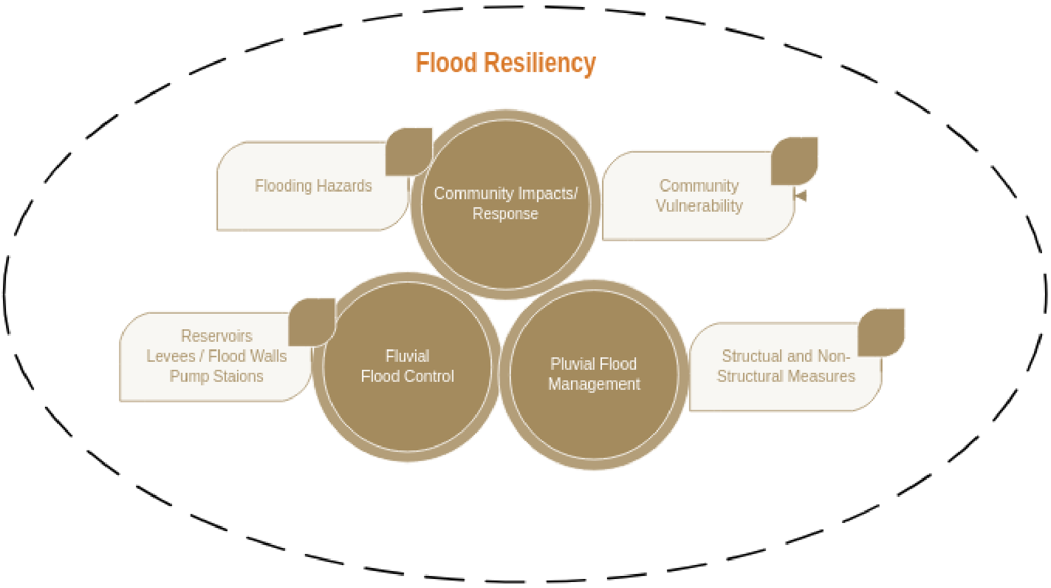

Figure 1. Resiliency Framework Adopted for Structuring the Review.

Physics-based methods and machine learning methods are not competitors and can be used in a synergistic manner. Therefore, a brief overview of some commonly used machine learning codes were briefly reviewed here. While machine learning methods have gained popularity and wide-spread usage, these tools and methods are not routinely taught in water related curricula. Many decision-makers may be unaware of different types of learning algorithms that Machine learning offers. Therefore a brief overview of machine learning methods are also reviewed for the sake of completeness. A narrative review approach was adopted for these topics as the goal was to summarize key concepts and provide context and structure for a more comprehensive machine learning models across the flooding life-cycle [16].

Having presented the conceptual life-cycle of flooding, a detailed systematic review of literature pertaining to each topic was undertaken [17]. Major bibliographic databases - 1) Web of Science –

 doi:10.20944/preprints202504.1887.v1

Core Collecton 2) Scopus and 3) Google Scholar were utilized to identify over 5,000 references using a suite of key words such as - 'flood reservoir inflows+ Machine learning"' "Flood reservoir outlfows + Machine Learning"; "Flood Resiliency - Machine learning". The search was constrained to a custom range of 2019 - 2025 (both inclusive) as the number of publications focused on machine learning have increased exponentially over this period, with many new advancements. The literature was trimmed down to those presented mostly in Q1, Q2 journals or in respectable conference proceedings such as IEEE conferences which undergo peer-review. The inclusion of conference proceedings were deemed necessary as some newer innovations particularly using sensors and ML appeared in such publications. Other journals were utilized when the paper had something unique to offer. The focus largely was on those publications that provided methodological advancements, rather than those which simply presented newer case-studies without any significant methodological contributions.

Given the vast and growing literature in this area, no claims of completeness are being made here. However, the review comprehensively captures major strands of research inquiry. The primary focus of this systematic review was to identify common themes and patterns in literature. Figure 1 presents the definition of resiliency and major themes around which the review is organized. Drawing insights with regards to ML algorithms used and factors influencing their performance. As performance metrics tend to be site specific and affected by the quality of data, the focus was not on quantitative evaluations but on how the models performed overall, what challenges were faced across different aspects of the flood management life cycle and what the unexplored directions were. The focus was on identifying themes and creating generalization framework that will enable decision-makers, researchers and practitioners evaluate future studies using the frameworks and insights presented here.

3. Evolution of Flood Management Practices

3.1. Fluvial Flood Control

Early flood mitigation efforts were largely focused on controlling the harmful effects of flooding to humans and was based on the notion that flooding, while inevitable, could be fully controlled. Flood control projects were largely structural in nature and involved building large-scale facilities to store flood waters and both delay and attenuate the flood wave to protect downstream communities. Even today, riverine flood control structures provide the first line of defense against flooding waters arising from upstream areas. These structures include reservoirs, levees, flood walls and pump stations.

Reservoirs are constructed across a river and usually have extra free-board and sluice-and-gate mechanisms to control flooding. Levees are earthern embankments built alongside the river to prevent water from overflowing. River walls serve the same function as levees but are often built with concrete. Pump stations are used to pump excess water that is flowing out of the levees or flood walls and pump that water back into the river.

Despite several flood control projects being undertaken, flooding continues to be a problem in many parts of the world [18]. Building new or scaling up existing flood control infrastructure is cost prohibitive and highly dependent upon the quality of hydrologic information that is available at the site [19,20]. The increased climate variability also adds uncertainty with regards to hydrological information and makes scale up of existing projects challenging [21].

3.2. Pluvial Flood Management

With increased urbanization, the impacts of pluvial flooding became significant. The growing recognition that controlling fluvial flooding was alone insufficient and humans and the environment must be protected from pluvial flooding. Flood management now sought to create an environment where humans can learn to live with floods. Retention and detention basins, culverts and green infrastructure such as rain gardens, bioswales, blue roofs and enhanced infiltration systems (e.g., permeable pavement) prevent flood water from accumulating in residential areas [22,23]. In a similar vein, non-structural measures such as zoning laws that restrict development in flood prone areas, increase public awareness and employ early-warning systems to evacuate people to safer areas are also

 doi:10.20944/preprints202504.1887.v1

utilized [24–26]. Rather than keeping the flood waters away from humans (as is the case with flood control), flood management also sought to keep humans away from flooding hazards. This approach of flood management was largely based on the notion of identifying the susceptibility of a land parcel to flooding (e.g., 100 year floodplain) and using that information to keep people away from flooding impacts through structural and non-structural measures[27].

While coastal and riparian areas are prone to flooding, they also offer several other significant opportunities such as access to water-based trade routes, fishing and aquaculture operations and enhanced recreation activities. The number of people living in coastal areas have increased substantially over the last few decades with nearly 40% of the world's population residing within 100 km of the coast [28] and over 50% of the population within a few kilometers of a water body [29]. Flood insurance programs provide safety net for people susceptible to flooding damages. These programs rely on risk assessments to evaluate the probability of a flood hazard that a person and/or infrastructure is exposed to in developing insurance policies and setting premiums.

While flood risks have been defined in multiple ways, using both qualitative and quantitative measures, widely used definition views flooding as a product of hazard, value and vulnerability [30]. Hazard characterizes the threatening event including its probability of occurrence, the values or (values at risk) denote humans, infrastructure and other assets that present at a location and vulnerability is the lack of resistance to withstand the damaging forces of the hazard causing event. Flood risk management uses the risks of flooding hazards as the basis for management and seeks to develop structural and non-structural measures that mitigate the effects of hazards to acceptable levels of risks [25].

3.3. Resiliency-Based Flood Management

While floods do cause damages to both humans and the environment, they are important for sustaining river and floodplain ecosystems. Floods transport fine sediments and energy required for riverine flora and fauna. In addition, the disturbances created by flood and droughts are essential for fostering ecological biodiversity [31,32]. The ecological importance of floods is no longer a theoretical construct but has been integrated with reservoir operations for flood control [33]. Flood management is transitioning from human-centric approaches towards more holistic frameworks [34].

The recognition of the importance of flooding for the functioning and sustainability of vibrant riparian ecosystems coupled with the inadequacy of flood control strategies, the movement of flood mitigation from being a command-control type engineering endeavor to a more stakeholder-driven, participatory community-based effort has brought to light the concept of flood resiliency. While the definitions of flood resiliency continue to evolve [35], the synthesis of resiliency definitions from a wide range of disciplines provided by [36] has seen widespread usage in resiliency-based flood management studies [37,38]. In this approach, flood resiliency approaches are categorized into three types - 1) Engineering resiliency; 2) Ecological resiliency and 3) Socio-ecological resiliency (see Figure 2). While the words engineering, ecological and socio-ecological are used in this approach, they are not domain-specific and have been used in diverse systems and applications [37].

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 April 2025 doi:10.20944/preprints202504.1887.v1

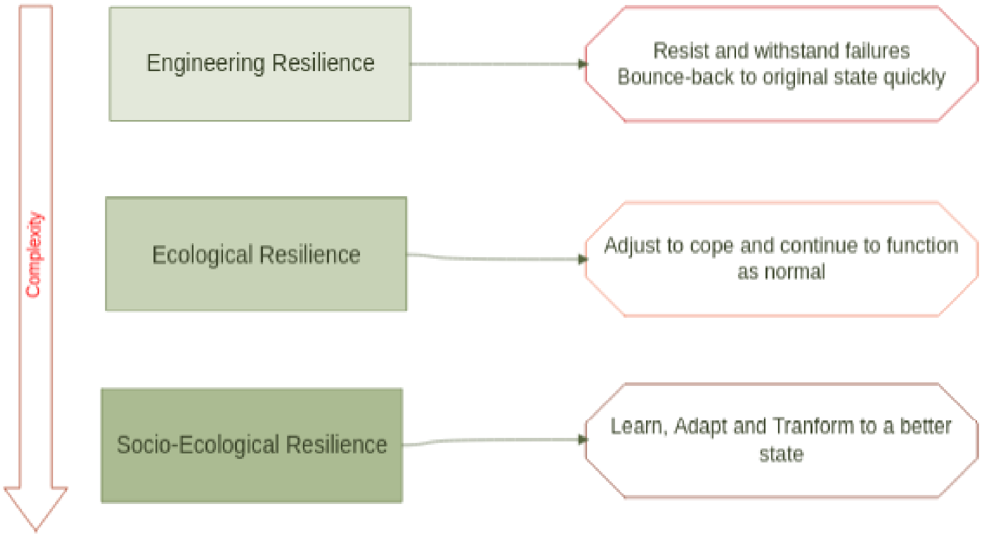

Figure 2. Common Definitions of Resiliency.

The main idea of engineering resilience is to create systems that can withstand hazards and not fail. If and when failure does occur, these systems need to bounce back to their original (pre-flooding) stage quickly. The traditional design of hydraulic infrastructure is based on the idea of engineering resiliency. The concept of return period (or the average time of recurrence of a flooding event of a certain magnitude or higher) is used for design of structures that aim to control floods (dams and levees), mitigate their effects (detention and retention basins) or otherwise interact with floods (e.g., bridges) and suitable for use in single systems [39].

The ecological resilience typically is applied to a system with multiple interconnecting parts. This allows the system to adjust and cope with flooding hazards. For example, early warning systems allow people to procure supplies in advance, take other precautionary measures (e.g., evacuate to a higher elevation) to cope with short-term floods. The socio-ecological resilience focuses on complex-adaptive systems (or a system of systems). Here the system may not come back to its original state but adapts and transforms to a better state. For example, relocation of people out of the floodplains to areas with lesser inundation risks are undertaken after many devastating hurricanes [40].

It is important to recognize that flood resiliency is not independent of flood control and flood risk management approaches. Rather, the resiliency paradigm subsumes both flood control and flood risk management concepts. From this viewpoint, the resiliency provided by flood control and risk-based structural and non-structural alternatives is evaluated and improved upon in most flood resiliency endeavors so as to benefit from past investments in flood planning and course correct in light of new information [38]. While the goal of any flood planning endeavor is to minimize the risks of flooding to humans and the environment, socio-ecological resiliency based planning and management challenges to think outside the box and evaluate if converging to a new equilibrium is better (or worse) than trying to get back to the original state.

Regardless of the methodology adopted, flood management is a costly endeavor that requires considerable investments in infrastructure and a time-consuming process where diverse and often conflicting views and values have to be reconciled effectively.

4. Physics-Based Models for Flood Management

Quantitative information is paramount to flood management. Estimates of runoff, streamflows, the nature and extent of inundation, the frequency and magnitudes of extreme rainfall are therefore a major component of flood management endeavors. It is not only important to analyze historical floods but also develop robust estimates for future conditions, especially given the recognition that past climate is not necessarily a good indicator of future flooding events [41]

Mathematical models used for flooding can be categorized as - 1) White-box models, 2) Grey-box models and 3) Black-box models. White-box models are most rigorous and employ mass, energy and momentum conservation principles. Grey-box models are based on the conservation of mass principle, along with empirical laws to describe flooding. 3) Black-box models are completely empirical and built on site-specific data.

Both white-box and grey-box models are referred to physics-based models as they are based on physical principles. The development and application of these models dates back to at least 100 years [42]. Physics-based models can be developed for a wide range of systems such as, streams and rivers, watersheds, reservoirs and wetlands. In addition, they can be developed to study combined interactions like - climate-hydrology dynamics [43], human-water linkages [44] and food-energy-water nexus [45].

Flood planning studies often employ more than one model. For example, a conceptual hydrological model can be used to model fluvial flood discharges into a stream, while a hydraulic routing model can be used to study fluvial flooding. When multiple models are used, they can be loosely-coupled (meaning the two models are separtely run) or tightly-coupled (one software allows running of all the models with seamless data transfer between them). The application of physics-based models entail many aspects. Salient modeling considerations are presented in Table 1

Table 1. Factors Affecting the Complexity of Models used in Flood Risk Management.

Concept	**Types**	**Remarks**
Model Spatial Dimensions	0D, 1D, 2D, 3D	0D models are also called lumped or box models
Model Spatial Discretization	Lumped, semi-distributed, fully-distributed	Semi-distributed models define the watershed using subwatersheds. Fully discretized models use grids. Hundreds to thousands of grid cells are used to cover the watershed of interest
Time Dimensions	Steady-state; Dynamic	Steady-state models are time-invariant, while dynamic models can vary in time (e.g., subhourly, hourly, daily, monthly, annually)
Event Type	Single Event, Continuous	Continuous models operate during wet and dry periods, while single event models simulate flooding associated with single rainfall event.
Process Description	Linear, nonlinear	A model is nonlinear if even one of the process is expressed using nonlinear equations
Solution Scheme	Analytical, Numerical	Analytical models use exact solutions, while numerical schemes use approximate methods such as the finite-element or finite-difference schemes

Several software simulators have been developed by various agencies to support flood management. Widely used software simulators are presented in Table 2. A wide range of engineered and

 doi:10.20944/preprints202504.1887.v1

natural systems pertaining to flood control can be modeled using these software at varying levels of complexity.

Table 2. Commonly used Software for Flood Management.

Software	Type	**Developer**	**Description**	**Major Outputs**
HEC-HMS	Lumped and Semi-Distributed	USACE	Pluvial flood forecasting; River routing	Outflow hydrographs; peak flow
HEC-RAS	Fully-Distributed (1D/2D)	USACE	River hydraulics, dam breach	Water elevation, inundation mapping; velocity
SWMM	Semi-Distributed	USEPA	Urban drainage, pluvial flooding, green infrastructure	Stormwater hydrograph, flood depth, sewer overflows
SWAT	Semi-Distributed	Texas Agrilife/USDA	Watershed streamflow, sediment and pollutant transport	Streamflow, flooding, long-term hydrology and water quality concentration
Mike 11	Lumped/Semi-Distributed	Delft Hydraulics Institute	1D River and channel modeling	Flood hydrograph, water level discharge
Mike 21 FM	Fully Distributed (2D)	Delft Hydraulics Institute	2D model for urban flooding	Water depth, velocity fields, flood inundation maps
Mike Flood	Fully Distributed (1D/2D)	Delft Hydraulics Institute	1D and 2D River and channel modeling. Integrates Mike 11 + Mike 21	Flood Hydrographs, Discharge, inundation maps
Mike Urban	Semi-Distributed	Delft Hydraulics Institute	Urban stormwater and pluvial flooding	Water levels, hydrograph and sewer outflows
TUFLOW	Fully Distributed (1D/2D/3D)	Tuflow.com	Stormwater, pluvial flooding, drainage networks	Water elevations, velocities, inundation extent
Flow-3D	Fully Distrbuted 3D	Flow Science Inc.	Computational Fluid Dynamics Model for dam break and complex urban flows	3D velocity profiles, water elevations and flood propagation

4.1. Shortcomings of Physics-based Models

While physics-based models offer great theoretical rigor, their implementation is challengng on several fronts. Physics-based models contain model inputs that cannot be directly measured. These parameters have to be estimated indirectly from available observations. The process of obtaining unknown model inputs using observed outputs is called the inverse-problem in mathematics or calibration in hydrologic literature. The inverse problem is mathematically ill-posed and therefore does not yield unique solutions. In other words, different subsets of model inputs can result in similar output predictions [46]. Typically available hydrologic records at most sites only allow reliable estimation of a few parameters [47]. Efforts to improve model calibration has focused on using expert knowledge to guide calibration [48] and/or regularization schemes that seek to restrict the number of calibration parameters during calibration [49,50], using other surrogate information [50] to guide calibration, using ensemble methods [51] to construct confidence bounds by employing parallel computing approaches [52]. However, these improvements still largely remain in the theoretical regime and not widely used in routine flood management studies. Sensitivity and uncertainty in model outputs due to uncertainty and variability of inputs is recommended to understand limitations of estimates like inundation obtained from physics-based models [53].

As physics-based models require calibration, their usage in ungaged basins becomes even more challenging due to lack of long-term output data [54]. Regionalization approaches are often employed to overcome this limitation. Calibrated data from nearby, and/or, similar systems are scaled for use in the ungaged basin [55,56]. Scaling of hydrological phenomena is however non-trival and the success of regionalization depends upon the amount of informative content that is present to facilitate scaling [57]. Nesting ungaged watersheds within gaged watersheds is beneficial if the gaging density of the larger watershed is sufficiently high [58].

The ability of physics-based models to capture flood peaks is also of concern. This limitation stems from several factors - 1) Most flood models operate on a time-step of days and months due to availability of rainfall records at those scales. However, flooding is a much more dynamic phenomenon that causes changes in flows in the order of minutes to hours. The coarse temporal discretization limits the ability of the model to capture fine-scale changes [59]. Models make several other approximations to simulate various hydrological processes, these simplifiations do well in capturing general trends, but cannot adopt to sudden changes in flow regimes [60]. 2) The Root mean square error (RMSE) or its variant, the Nash-Sutcliffe Efficiency (NSE) are widely used for model calibration. Minimization of this metric entails a trade-off in capturing a large number of low-flow events and relatively few high-flow events which often leads to over-estimation of low flows and under-estimation of peak flows. The use of newer error metrics like the Kling-Gupta Efficiency (KGE) [61] can ameliorate this impact [62]. 3) The amount of data available also impacts the abiliy capture peak flows [63]. Clearly, errors in flood forecasts induces uncertainty in flood planning and management.

5. Machine Learning Modeling - Background

The limitations associated with physics-based formulations coupled with recent and rapid advancements in artificial intelligence (AI) and machine learning (ML) has increased the exploration of machine learning methods in applications focused on flood management. Machine learning models are not new, the term was coined by Arthur Samuel back in the year 1959 [64]. A timeline of major advancements in machine learning is presented in Figure 3. Algorithms developed in the 1960s such as K-nearest neighbors (KNN) are in use even today [65].

Artificial Neural Networks (ANN) have been around at least since mid 1980s and some of the early application of this technique for streamflow forecasting dates back to early 1990s [66,67]. In a similar vein, an early application of tree-based learners for flood studies can be traced all the way back to 1992 [68]. The number of applications of machine learning models has shown steady increasing pace through 1990s - 2010s congruent with algorithmic developments in this period.

In recent years, the number of papers has grown exponentially given the large-scale availability of data, general purpose libraries such as Scikit Learn [69] and Tensorflow [70]. Computational power has also increased significantly in the last decade with the availability of graphical processing units (GPU) and Tensor Processing Units (TPU) on some platforms. Unlike conventional central processing units (CPU), GPU support parallel floating point operations (FLOPs) that speeds up large-scale computations. TPUs utilize a grid of parallel processing elements through which computations are processed in a synchronous manner.

doi:10.20944/preprints202504.1887.v1

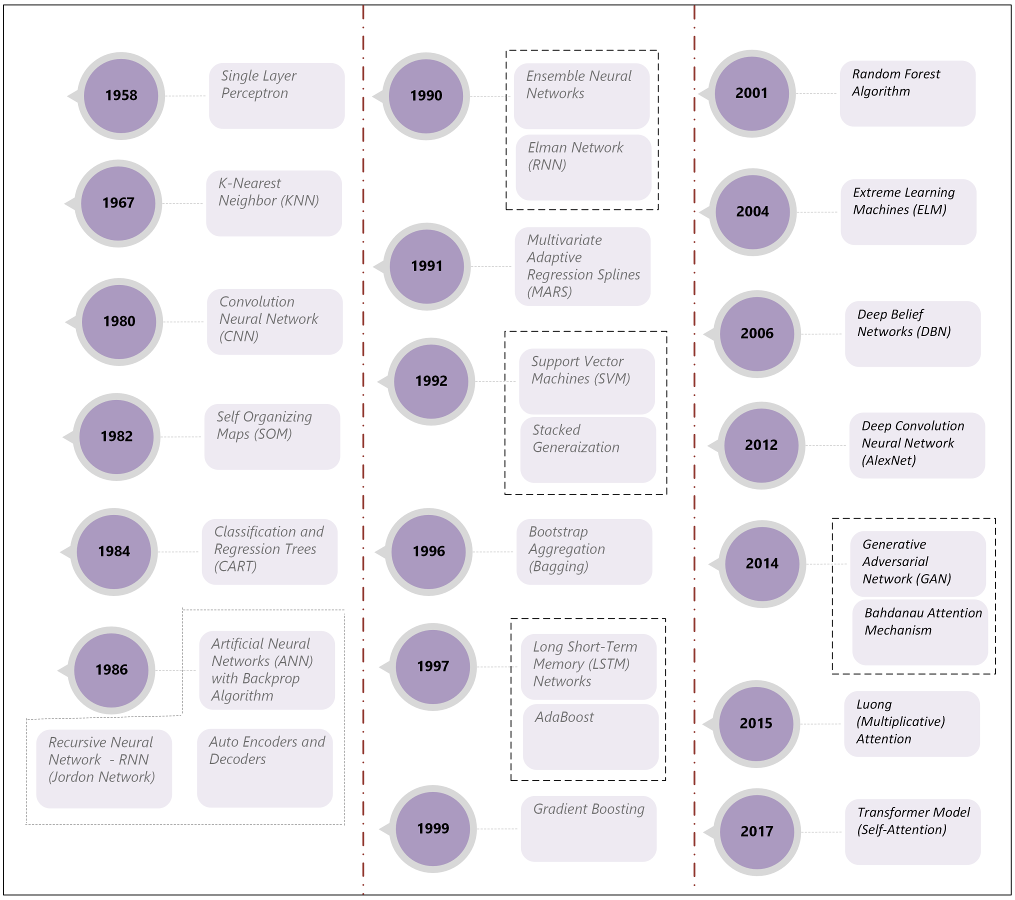

Figure 3. Timeline of Major Breakthroughs in Machine Learning

Prior to reviewing the use of machine learning for flood management, a brief overview of major learning strategies employed is presented to contextualize their utility in flood management studies.

5.1. Machine Learning Strategies

As shown in Figure 4, the four main types of ML are – supervised learning, semi-supervised learning, unsupervised learning and reinforcement learning. In supervised learning, a ML model is trained using a datasets that has both the inputs (features) and the output (labels). Outputs can be continuous or discrete. Supervised learning methods that employ discrete labels are referred to as classifiers, while those mapping continuous labels are called regressors. Unsupervised learning on the other hand is trained on a dataset that has no labels and is often used to cluster data. Semi-supervised learning uses a dataset where there are some data with labels and others without labels. This approach can be used for both classification and regression tasks. Reinforcement learning (RL) focuses on developing a long-term (operating) policy that maximizes long-term reward , while RL algorithms can in theory be used for classification, regression and clustering tasks, they are not specifically designed with these applications in mind, but are better suited for decision making and control [71].

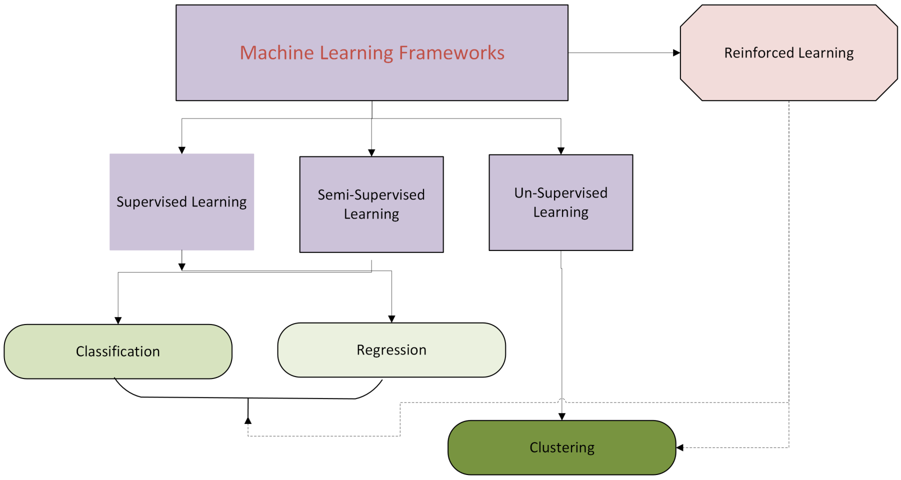

Figure 4. Timeline of Major Breakthroughs in Machine Learning

Humans use different approaches to learn information. Cognitive approaches focus on using past experiences and knowledge for making decisions and solving new problems. These strategies can range from simple memorization of facts, forming rules of thumb to guide decisions. In a similar vein, humans also decide when to learn what information, evaluate and discern among competing pieces of information to make a decision based on the weight of the evidence provided by them. Human decision making is not always individualistic but involves collaboration (bagging) with peers and partners. Humans also exhibit the ability to build and improve on their previous experiences (boosting). Machine learning models that try and mimic the cognitive capabilities of humans can be categorized as cognitive-inspired machine learning (CIML) approaches.

Biological adaptations modify how stimuli (data) are sent to the brain (central processing unit) and how they are processed to create new information. Evolutionary changes happen over longer time-scale and change the biological makeup to improve survival. On shorter timescales, humans (living beings) also exhibit the ability to change their physical and neurological conditions to adapt to new conditions, learn new skills or adapt to a new environment. In particular, the brain can reorganize and form new neural connections – this phenomenon is referred to as neuroplasticity and a key mechanism by which the brain adapts to new information and environments.

Neuroplasticity plays a key role in memory, mortar learning and helps recover from brain injuries and is controlled by the nature of the stimuli [72] Machine learning models that mimic evolution and/or the structure of the nervous system and the functioning of the brain can be categorized as biologically-inspired machine learning models (BIML).

A variety of machine learning models have been developed over the last seven decades (see Figure 4). Learning strategies provide a quick way to understand the broader ideas related to machine learning algorithms. Table 3 provides a grouping of machine learning algorithms based on their learning styles.

 doi:10.20944/preprints202504.1887.v1

Table 3. Learning Strategies of Machine Learning Algorithms.

Learning Strategy	Learning Description	Style	Machine Learning Type	Example Method
Rule-based learners (A type of associative learning)	Codifies the relationship as IF-THEN rules	CIML	Supervised	Classification and Regression Trees (CART), Multivariate Adaptive Regression Splines (MARS)
Lazy Learning	Memorizes data, defers learning until prediction time	CIML	Supervised	K-Nearest Neighbor (KNN); Case-based Reasoning (CBR)
Eager Learning	Learns general function during training	CIML & BIML	Supervised	Support Vector Machines (SVM); Naive-Bayes (NB), Artificial Neural Networks (ANN)
Reinforcement Learning	Learns by interacting with the environment with rewards and penalties	CIML	Reinforcement	Q-Learning, Policy Gradients
Evolutionary Learning	Learns using principles of genetic encoding and survival of the fittest paradigms	BIML	Supervised, unsupervised and reinforcement learning	Supervised – Symbolic regression; Unsupervised – Evolutionary Clustering;Reinforcement – symbolic policy learning
Hebbian Learning	Learns by strengtening co-activating neurons	BIML	Unsupervised learning	Self-organizing Maps (SOM); Neural Hebbian Nets
Corrective Learning (Delta rule learning)	Leans by minimizing errors of predictions	BIML	Supervised & Unsupervised learning	Artificial Neural Networks (ANN) using backpropogation or it variants
Greedy learning	Makes locally optimal decisions at each step rather than seeking a globally optimal solution	CIML and BIML	Supervised Learning	Approach adopted by CART, deep neural nets and other algorithms that have a large of decision variables

 doi:10.20944/preprints202504.1887.v1

Competitive learning	Competition helps select a winner	CIML and BIML	Unsupervised & Supervised learning	Self-Organizing Maps (SOM); Ensemble classifiers
Active learning	Model queries a human or a database for informative samples while learning	CIML and BIML	Mostly supervised learning	Diversity Sampling (DS)and Query-by-Committee (QBC), uncertainty determination and minimization
Multi-task learning	Model learns more than one output from a set of inputs	CIML	Supervised	The idea of multi-tasking is cognitive, but BIML models can be used to achive this cognition such as multi-input-multi output (MIMO) models - MIMO-ANN
Collaborative Learning	Multiple models are developed and combined to improve predictions	CIML and BIML	Unsupervised and supervised	CIML and BIML models can be combined in this approach
Bagging	Trains different models using bootstrapped samples of data	CIML	Supervised	Random Forest, Bagging Trees – A Special form of Ensemble learning
Boosting	Trains sequential models to correct previous errors	CIML	Supervised	AdaBoost, Gradient Boost, XG Boost. Another form of Ensemble learning
Deep Learning	Typically a neural network model with multiple layers to handle big data	BIML	Supervised and Unsupervised	Deep Belief Networks and many variants
Attention Learning	Focuses on most important data for prediction	BIML	Supervised	Commonly used in deep neural networks
Adversarial Learning	Generator and Discriminator compete to improve model. Akin to Predator-Prey dynamics	BIML	Traditionally unsupervised, but can be modified for supervised learning	Generative Adversarial Network (GAN); A deep learning method. Conditional GAN (cGAN) for supervised learning

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 April 2025 doi:10.20944/preprints202504.1887.v1

Recurrent Learning	Remembers previous data via memory cells or recurrent connections. Used with sequential data such as time-series and text sequences	BIML	Supervised learning	Long Short-Term Memory (LSTM) network, Elman Machines. A form of deep learning
Convolutional Learning	Uses moving windows to sample features, pool data to retain importan features and them perform nonlinear mapping	BIML	Supervised learning	Useful for gridded data. Convolution Neural Networks (CNN). A form of deep learning.
Encoder-Decoder Learning	Encodes and decodes data and useful for data compression, generation and transformation	BIML	Supervised and Unsupervised	Autoencoders-Decoders; Transformer models. A form of deep learning
Self-Supervised Learning	Model creates 'pseudo-labels' from input data for training supervised models	BIML	Unsupervised / Hybrid	Used to fill missing values particularly in images.

More than one learning strategy presented in Table 3 can be adopted for a given task such as mapping nonlinear input-output relationships.

5.2. Calibration of Machine Learing Models

Machine learning models are empirical in nature, therefore the model parameters are unknown and have to be determined via calibration, particularly in the case of supervised learning. In unsupervised learning, there is no calibration per se, but distance between points within a cluster are minimized while the distance between two independent clusters are simultaneously optimized. While physics-based models are often calibrated using both automatic (optimization-based) and manual methods, machine learning models are calibrated almost exclusively using automatic optimization based methods. In case of supervised classification, an error metric, such as the root-mean square error (RMSE) is minimized. The objective function that is minimized is commonly referred to as the loss function in machine learning literature. Similarly, the calibration of the supervised learning model is referred to as training the model.

Typically, the calibration of a supervised ML is carried out by splitting the available dataset into two parts. The first part is called the training dataset which is used to train the model and the second part the testing dataset. The training dataset is used to obtain unknown model parameters (e.g., weights of neural network) by minimizing a loss function. The performance of a model is then assessed by comparing the model predictions against the testing dataset that contains data that the model has not seen before. ML model algorithms also have a set of parameters that are necessary to define the structure of the model or to implement the optimization routines necessary for model calibration. These parameters are called 'hyper-parameters'.

Hyper-parameters are either set based on previous studies, but most often are estimated (optimized) during the calibration process. When hyper-parameters are optimized as part of the training process, the training dataset is further divided into two parts – Training Data and Validation dataset. The training dataset is used to train the model, the validation data is initially used to test the performance of the training under a given set of hyper-parameters and select their optimal values. Once the hyper-parameter values are identified, the validation dataset is subsumed into training dataset and a final training is performed on the model. Validation datasets are also useful for early stopping (i.e., when the validation error increases even when the training error is decreasing as training progresses. Similarly, the validation dataset can be used to select the best model for a subset of models or use the validation error metrics to weigh different models within an ensemble.

Typically, the dataset is randomly split into training and testing subsets. In case of sequential data (e.g., time-series) the first part of the data is used for training while the latter part of the data is kept aside for testing. Train-testing splits of 70%-30% or 80% - 20% are common. Typically, 10% - 20% of the training data is used for validation.

Overfitting occurs when a model is able to predict the training data very well, but demonstrates a rather poor performance on the testing (independent) dataset. Overfitting generally implies the model has too many degrees of freedom(parameters) than necessary. Thus the model is able to learn the noise in the data, rather than focus on generalization. Overfitting can also occur with hyper-parameters. The validation dataset is used to avoid overfitting of hyperparameters. Regularization methods are widely used to prevent over-fitting, a regularization function adds a penalty with increasing weights, which causes the model to effectively drop some weights during the calibration to reduce the penalty on model performance. Overfitting is more likely in deep neural network models due to their complexity and more aggressive approaches such as dropout (where certain nodes are dropped from training with a certain probability) are used to prevent over-fitting.

Training of supervised learning models typically occur by sending small batches of data through the model to adjust weights. Each run with a model is called a training epoch. The stochastic gradient descent algorithm (SGD) is commonly used for model calibration [73]. Automatic or Algorithmic differentiation (AD) is widely used to compute gradients without significant numerical errors [74]. These gradients are used to identify newer estimates of parameters which reduce the loss function to the lowest possible value. However, in deep neural networks, the gradients of weights of early layers practically become zero (i.e., vanish) making the training using conventional gradient descent ineffective. Greedy learning uses layer-by-layer training to eliminate the effects of vanishing gradients and effectively training deep networks. Greedy learning approaches are also used in other ML models such as random forests and gradient boosting methods [75].

5.3. Explainable Machine Learning

While machine learning models can provide accurate predictions and map highly nonlinear input-output relationships, interpretability of the model is a major limitation. In this regard, physics-based models and statistical models (e.g., linear regression) perform much better as the model equations and parameters can be readily interpreted.

Explainable machine learning (XAI) is a rapidly emerging field of ML that seeks to improve the understanding of the inner workings of ML models. XAI provides several benefits, most important of which is the increased trust in using this models in mission critical applications such as flood forecasting. XAI also helps evaluate if the model is based on biased data and help debug the models and improve their work.

Some AI techniques such as CART, MARS and Genetic Programming (GP) encode information in the data as rules and equations and therefore are readily interpretable. As boosting and bagging methods create multiple models with the same dataset, the relative importance of different parameters can be ascertained even if the model equations are not explictly known per se. Tree-based algorithms are also used to explain other hard to interpret models [76]

Neural network models (particularly deep learners) encode the information in terms of weights that are hard to interpret. In such cases, the outputs obtained from the model are explained by fitting local (explanable) models around the predictions – The Shapley Additive Explanations (SHAP) that uses game theoretic measures to explain the importance of different features on the output and the Local Interpretable Model-Agnostic Explanations (LIME) based on linear regression around the predictions are commonly used to explain which inputs are important and how they affect a particular model output [77].

Having presented a broad overview of machine learning methods, the application of these techniques in flood risk management are explored next.

6. Machine Learning for Flood Resiliency

6.1. Machine Learning for Fluvial Flood Control

6.1.1. Machine Learning for Reservoir Operations

Reservoirs store excess water during high flow events and release them during periods of high demands. The controlled release of flow out of the reservoir is to ensure there is minimal downstream hazards, but also is constrained by the capacity of the reservoir and associated infrastructure. Therefore, the outflow of the reservoir is a function of the water height in it. Reservoir operation rules (ROR) are used to define outflows. Operation rules tend to be complex, especially when there are multiple interconnected reservoirs, and they are only known to a few practitioners.

Machine learning models have been used to infer reservoir operating rules and predict outflows. A wide range of machine learning approaches including artificial neural networks (ANN), Support vector machines (SVM), random forest (RF), deep neural networks (DNN) and recurrent networks, especially the long short-term memory (LSTM) network have been adopted for this purpose [78–85].

The main synthesis of these studies is that machine learning learning models provide better predictions than conventional tools (e.g., statistical models such as ARIMA or water balance formulations). Deep learners, especially LSTM and GRU (gated recurrent units) provide better predictions than other techniques in most, if not all, cases. However, there is no universal consensus on which model always provides the best results. Different models may perform well at different sites within a river basin. Therefore, assessment must always include multiple models to select the most optimal of those tested.

Reservoir inflows are key input to these models and the quality of the output depends upon the quality of the inflow data. Note reservoir inputs can be natural streamflows or controlled releases from an upstream reservoir (these reservoirs may or may not be explicitly modeled). While treating upstream reservoir releases as inputs simplifies the outflow predictions at downstream reservoir, explicitly modeling the releases of these upstream reservoirs is shown to improve model predictions due to better accounting of lag-times of flows [86].

Most studies focus on modeling monthly or daily outflows although efforts are beginning to be made to develop sub-daily streamflow time-series to improve streamflow forecasts [87]. Most models provide 1-step ahead forecasting with many providing 2-, 3- and multi-step ahead forecasts. Real-time forecasting is also being attempted using machine learning [88].

Statistical metrics such as the root mean square error (RMSE) are used as the objective (loss) function to minimize the errors between observed and modeled outflows. However, recent efforts have focused on including other information to help guide the calibration process. A mass conserving LSTM (mc-LSTM) model has been presented as well [89]. Loss functions have been modified to include conservation relationships as well as other information such as reservoir operation rules [85,90]

While recurrent and non-recurrent neural networks (e.g., LSTM, GRU and ANN) appear to provide better forecasts than most other models, the derived operational rules are abstracted as weights and thus not readily interpretable. This issue if not of concern if forecasting is the only goal of the model. However, interpreting the model results and which inputs have bigger impact on the output are useful to evaluate the performance of ML models (esp. those that behave like black boxes).

 doi:10.20944/preprints202504.1887.v1

Scenario-based sensitivity analysis, and XAI methods like the Shapely Additive Explanations (SHAP) are being employed to understand how ML model inputs affect reservoir outflows[90,91]. Results from these studies suggest that while XAI results make hydrological sense in most situations, there is no guarantee of this over the entire range of outputs [91]. An evaluation of whether ML models trained with physics-based or operation-based loss functions explain the results better is an interesting, but unexplored direction of research.

Reservoir inflows are critical to properly predict outflows. In forecasting applications, reservoir inflows have to be estimated to make future predictive runs.

Therefore, considerable interest also exists in using machine learning for modeling reservoir inflows [92–108]. These models mainly focus on daily time-step, but both monthly and sub-daily time-steps are also predicted depending upon the needs of the problem.

The synthesis of reservoir inflows using machine learing bring about several distinct insights - 1) These models use meteorological variables such as precipitation, temperature, wind speed and their lags. 2) Atmospheric teleconnections such as El-Nino-Southern Oscillations (ENSO); Noth Atlantic Oscillations (NAO), Pacific Decadal Oscillation (PDO), Indian Ocean Dipole (IOD) and Arctic Oscillation (AO) are also being employed. These teleconnections correlate with several reservoir inflow processes such as rainfall, evapotranspiration, and snowmelt. 3) There is a greater propensity to combine multiple models, rather than rely on a single algorithm such as LSTM. 4) Rather than utilize an ensemble mean (EM) most studies utilize stacking ensemble (SE) wherein different parts of the dataset are predicted by different model subsets. This occurs because, some models provide very high extreme values which bias the overall ensemble mean. 5) It is more common to see fused traditional and deep machine learning algorithms as well as statistical models such as ARIMA for reservoir inflow forecasts.

Reservoir inflow estimation is often fraught with uncertainty. These uncertainties can be random due to changing weather pattern, land use alterations as well as climate change. The ability of the model to capture extremely high inflows is critical in flood management applications. While ML models may perform better than traditional physics-based models for forecasting streamflows, they are not perfect nor without limitations. Model stacking and ensemble approaches may help with predictions, but also can potentially average down extreme forecasts. Explicitly capturing this uncertainty alongside point estimates adds value to the estimation process as it informs the decision maker of what is more likely estiamates and what are all plausible values. Therefore, many studies are exploring the use of uncertainty quantification methods such as Bayesian Deep learning, Bayesian networks, fuzzy logic and other heuristic models [48,92,106,107,109,110].

Despite growing number of publications in this area, propagating uncertainties in reservoir inflows through single and multi-reservoir operations is clearly a largely unexplored direction and of considerable practical significance.

6.1.2. Levees and Floodwalls

Levees and flood walls are structures built along the river to keep flood waters from flowing into the adjoining floodplain. While levees and flood walls serve similar purposes, the former is built using earthen materials, while the latter is built using concrete.

While applications of machine learning algorithms to study levees have been limited (see Table 4, they have tackled most challenging problems related to the resiliency of levees to withstand failure. Machine learning has been used to evaluate failure risk due to overtopping, compaction and liquifaction and piping. Cracks and other early warning signals such as the formation of sand boils which indicate piping due to liquifaction have also been studied. Manual inspection and ground penetrating radar (GPR) are two common approaches to evaluate hazards in these systems. Both of them are expensive and time-consuming, therefore automated detection using drones and use of transfer learning algorithms to enhance the value of limited GPR based surveys have also been explored using machine learning methods. Deep neural networks such as Convolution Neural Networks (CNN) and Viola-Jones object detection method which is based on the boosting technique have shown promise for evaluating drone-based surveys. Conventional neural networks such as ANN, SVM, Naive Bayes and others have been useful to predict risk of failures.

Table 4. Machine Learning Modeling focused on Levees.

Focus	Method	Data	Reference
Levee Overtopping	Logistic Regression	Geometric, Hydraulic Geotechnical	[111]
Levee Anomalies	AdaBoost ; Viola-Jones Detector	Field Inspection Data	[112]
Failure Hazards	Deep Learner	Electrical Resistivity Data	[113]
Hazard Classification	Clustering	UAV, Geophysical (shear velocity, EMI, Apparent Resistivity)	[114]
Levee Compaction	Deep transfer learning, ANN, KNN, NB, LR for prediction	Transfer Learning for Feature Dataset	[115]
Sand Boils	Stack of ML Algorithms (SVM. ANN, CNN)	Field Surveys (Images)	[116]

The United States alone has over 24,000 miles of Levees and the average age of these levees is 51 years [117]. As such, many of the levees are in need of rehabilitation and most levees were not designed for modern climatic conditions. The review of literature indicates that while attempts have been made to evaluate levee failure risks using machine learning, this is another largely unexplored area as far as flood resiliency is concerned.

The National Levee Database (NLD, 2025) [117] provides the best publicly available information to understand levee performance. Efforts to compile a larger dataset of images pertaining to levee failures, along with hydraulic, geotechnical and geophysical characteristics would be a useful addition to facilitate further research in this area.

6.1.3. Pumping Stations for Flood Control

Pumping stations offer the third line of defense in flood control. Pumps are used to remove any excessive water and pump it back into the river. Machine learning has been utilized to model this system. Lee and Lee (2022) utilized multi-layer perceptrons (ANN) to simulate the discharge from pumping stations using previous rainfall and discharges [118]. They conclude that while the method was useful, the study could have benefited from having inflow data into the pump stations as one of the inputs. Wang et al.,(2023) applied ANN, SVM with different evolutionary optimization techniques to study inflows into a pumping station along different time periods. They made use of water levels at other pumping stations, rainfall and human factors to develop different models that showed a high degree of accuracy in their study area [119]. Kow et al., (2024) [120] utilized transformer based LSTM (T-LSTM) to abstract pump station operating rules using rainfall and flows from upstream stations. The use of transformers assisted in capturing the seasonality in the operations and provided excellent results both for typhoon and convective storm events. Joo et al., (2024) [121] combine gated recurrent units (GRU) with deep-Q reinforcement learning to simultaneously minimize water levels in the retention basin and the number of pump switches over the entire storm duration. Their results indicate that this combination is useful to improve operations of pump networks used in flood control.

While traditional methods for operating pump stations using physics-informed models and traditional optimization routines continue to be in use [122], efforts have started on exploring the use

 doi:10.20944/preprints202504.1887.v1

of machine learning models for predicting inflows, outflows and operation of these pumping stations. Pump stations help reduce flooding risks to humans and the environment but can also cause changes to local hydrology as well as affect riparian and riverine habitats due to changes in dissolved oxygen [123]. Therefore proper operations of these systems is critical. While the use of machine learning shows great promise, the lack of open-source datasets, limits the number of researchers who have access to such data, which in turn hinders a deeper exploration of this topic.

6.2. Maching Learning for Pluvial Flood Management

6.2.1. Pluvial Flood Estimation

Machine learning methods have been widely used in pluvial flooding studies [124–145]. Figure 5 explains the primary goals of these machine learning models in pluvial flooding studies. These studies either focus on predictions of flood characteristics or susceptibility mapping of flood prone areas, mostly in urban environments.

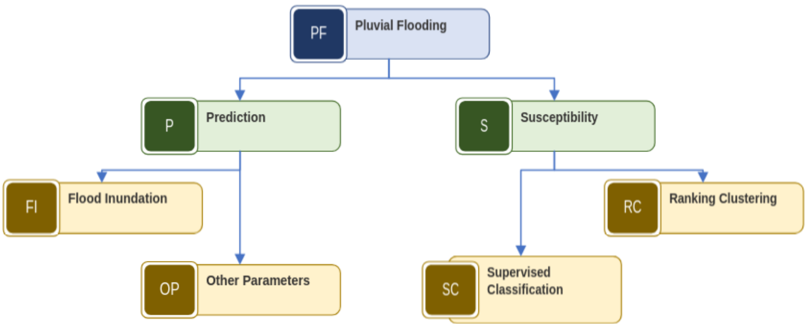

Figure 5. Classification of Pluvial Flooding Studies

The analysis of the literature pertaining to the use of machine learning for pluvial flooding demonstrates that a wide range of machine learning methods have been used in various applications. Among the traditional machine learning methods logistic regression (LR), artificial neural networks (ANN), support vector machines (SVM) and k-nearest neighbor (KNN) algorithms show greater use. Ensemble methods such as random forest (RF) and XGBoost have also been used extensively. Among the deep learning algorithms, convolution neural networks (CNN) stands out prominently. While all these models provide good results, there is no clear identification of which one is best.

The use of supervised ML methods in pluvial flood forecasting can be broadly categorized into two groups - 1) Classification methods are used to predict flood/no-flood zones using a suite of inputs such as precipitation, land use type, elevation and others. The labels (output) is obtained either from observed data within an urban area that is either collected by municipalities and/or those reported by citizens through navigation apps such as WAZE. The latter is useful for understanding flood prone roads and other transportation infrastructure.

Unsupervised methods, such as fuzzy clustering, have been used to categorize an area's susceptibility to flooding. Such classifications can help data collection activities to develop supervised classifiers in the future. The depth of the water level in urban areas following a rainfall event as well as the extent of inundation in floodplains are also important aspects of pluvial flooding. As urban areas are typically ungaged, this information has to be obtained via simulation or using satellite remote sensing [146–148].

Machine learning methods have been widely used to abstract the results from physics-based models (i.e., flooding depth, inundation extent) using surrogate measures such as rainfall intensity, land elevation and other characteristics. Again, a suite of conventional and deep learners have been employed for this purpose that have yielded good results. Once trained, machine learning

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 April 2025 doi:10.20944/preprints202504.1887.v1

models are easier to run and different scenarios can be evaluated to guide flooding planning and emergency management efforts. However, as ML models depend upon data from physics-based models, their accuracy cannot be greater than those obtained via physics-based simulators. Therefore, integrating physics-based simulations with satellite data can improve prediction methods, especially using interferometric synthetic aperture radar (InSAR) and other remote-sensing products. A case-study of such integration is presented in [148].

One important aspect of pluvial flooding is the heterogeneity of datasets. For example, rainfall hyetographs are 1D temporal data, while digital elevation models are 2D (raster data), high water marks are points in 2D space (vector data). Simulation models can produce data in 1D or 2D in space and over time. Transportation infrastructure is usually multi-line data (line vectors). Assimilation of such data from diverse sources is an important ancillary topic that is beginning to be explored [149,150].

Estimation of pluvial flooding characteristics, particularly inundation depth and extent using physically-based tools is rather difficult to paucity of data for model validation. While several studies have demonstrated that the output of such models can be captured using machine learning methods, the ability to transfer such learning to other sites via transfer learning approaches is another interesting area that has received little attention [151] and offers significant opportunities to improve flood risk management in other areas. In a similar vein, while the use of explainable artificial intelligence (XAI) methods such as SHAP have been demonstrated in the literature [134], their usage is still limited and offers a significant potential to eludicate factors affecting pluvial flood risks.

6.2.2. Machine Learning Approaches for Low Impact Development

Low impact development (LID) aims to develop stormwater infrastructure within urban areas that largely mimic natural hydrology by enhancing infiltration, supporting vegetation and enhancing water quality. In many US cities, they are part of the Municipal Separate Storm Water Systems (MS4) and permitted to discharge water from urban areas to natural water bodies. Environmental permits not only emphasize the rate and volume of discharges but also place restrictions on the quality of the discharging water. Therefore, contaminants of concern (CoC) often control their design and operation even if their primary goal is to remove flood waters from roads, residential neighborhoods and other business districts and keep pluvial flood waters away from humans.

Machine learning models are being used to design and operate many LID features and understand their functioning. Complex physics-based formulation can be approximated using machine learning and used with multi-objective optimization for sizing LID practices[152]. This approach greatly reduces the computational time to obtain good optimal solutions.

Deep reinforcement learning are widely being used to operate (control) stormwater treatment systems to mitigate flood flowrates and associated contaminant movement. Reinforcement learning is noted to help achieve real-time control under uncertainty [153–155].

Integration of machine learning with other ancillary technologies such as the Internet of Things (IoT), physical-model simulations and remotely-sensed datasets open new avenues to model the performance of stormwater systems, develop better designs and monitor LID systems [156,157]. Emerging studies indicate the promise of such integrations and this is another fertile area for further exploring the role of machine learning.

Predictions of water quality outflows from LID stormwater systems is of interest to decision makers to ensure they are within compliance. Machine learning models have found uses in capturing and correlating highly complex water quality transformations occurring within LID to easily measured surrogate data such as rainfall, temperature, sunlight and others. A variety of modeling algorithms have been evaluated and demonstrated to provide good results. Algorithms tested include, but are not limited to, ANN, SVM, KNN, Random Forest, Generalized linear models (GLM), Partial least squares (PLS) and deep learners [158–162]

While ML models show high potential to predict water quantity and quality parameters in stormwater control and treatment systems, there is no single approach that stands out. The quality of predictions varies among models so experimentation with a candidate set of models is often required. Conventional machine learning models (ANN, SVM) do better than deep learners in some cases, especially when the amount of data available for calibration and testing is small. In addition to variability across models, a particular algorithm my be better suited for one water quality parameter but not another. This result indicates that some ML schemes are able to better assimilate certain fate processes than others that affect the quality of stormwater. However, the success of ML models to predict emerging contaminants such as microplastics, indicates their high potential to predict contaminant concentrations whose mechanisms of biochemical transformations are not completely known [162].

While ML algorithms show promise for modeling urban stormwater infrastructure, there are many unexplored research directions. The use of explainable machine learning (XAI) to understand the predictions of ML models, especially to understand fate and transport of emerging contaminants has potential to develop fundamental insights on pollutant behavior using machine learning. Evaluating the use of stacked and ensemble learning methods, and integrating physics-informed loss functions are some easier implementations that have not been undertaken to-date to the best of the authors knowledge and offer some useful avenues to explore.

6.3. Machine Learning and Flood Resiliency

A significant amount of research focused on the resiliency of communities to flooding risks is beginning to emerge in the last few years and a large sampling of such studies, particularly those using machine learning approaches are summarized in Table 5. A synthesis of these studies depicts a common theme which is presented in Figure 6.

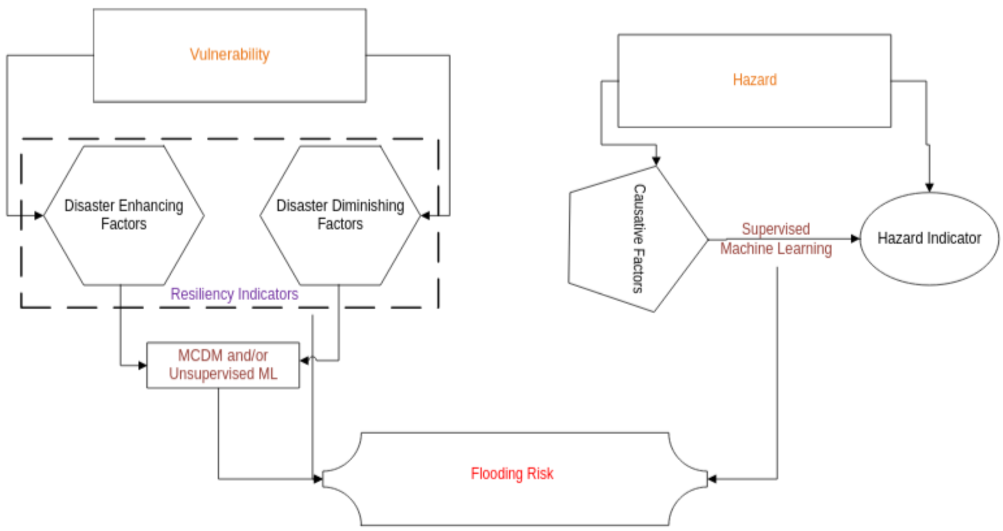

Figure 6. Overview of Resiliency-Risk Modeling Framework

Resiliency is measured as the extent of vulnerability of the community to withstand flooding hazard. Higher the vulnerability, the lower the resiliency of the community. Vulnerability captures the demographic, economic, geographic and engineering factors such as the population density, the nature and extent of economic activity, land elevation and distance from flooding features such as streams and engineering measures such as distance to pumping stations and other stormwater conveyance features.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 April 2025 doi:10.20944/preprints202504.1887.v1

Table 5. Summary of Studies Focused on Flooding Risk and Resilience

Citation	Methods Used	Focus / Findings
[163]	AHP, GIS, RS and ML (Random Forest and SVM)	Community resilience of floods; ML for susceptibility analysis
[164]	CNN to extract features for flood resilience + Fuzzy logic for Resilience Index	Resistance, Functional and Economic resilience explicitly modeled
[165]	SOM compared with PCA	Economic, Physical, Social dimensions of vulnerability considered.
[166]	SVM, ANN, RF, GBDT; Stacking and Ensemble not superior	Meteorological, Geographic and Human Resilience explicitly considered.
[167]	Unsupervised, Supervised	Categorize resilience and predict using rainfall
[168]	Different clustering;	Resilience defined in terms of - Robustness and Rapidity
[169]	NB, LR, RF, Lazy Tree, ANN (RF provided the best results)	Predict 4 classes - Flood, Flash Flood, Coastal Flood and Lakeshore Flood
[170]	Ensemble methods, Random Forest	Flood risk is Product Flood susceptibility based on weights of evidence and flood hazard maps
[171]	SVM, XGBoost, RF, MLP GBDT, 1DCNN	Disaster inducing factor, disaster breeding element, disaster bearing bodies (Input); Deeper models not as useful as shallower models.
[172]	SVM + MCDM	SVM for flood susceptibility; MCDM for flood vulnerability
[173]	Random forest was the best; SVM and Boosted Regression Trees	TOPSIS for Vulnerability; ML for hazard; Output - Various levels of risk
[174]	Tree-based approaches with DEA	DEA for Integrating Socio-Economic and Adaptive Capacity indicators (Vulnerability); ML with Geomorphology for Susceptibility
[175]	XGB, RF, CatBoost - RF was the best model	Susceptibility in riverbeds; Geomorphology for mapping susceptibility.
[176]	GARP and MaxEnt	Hazard (Flood, Economic, Social) x Hazard (Flood hazard - Based on Field Survey and ML)
[177]	ANN and linear Regression	Landscape factors affecting Flood Susceptibility (Satellite for flood hazard; Role of LULC via Regression)
[178]	Text Mining; Clustering and Prediction	Role of Big Data, IoT and Social Media data - Flood risk mapping; Rapid Impact Assessment on Infrastructure failure; Smart Situational Awareness; More conceptual study with an application to Harris County.
[179]	Random Forest and SVM; Multiple MCDM methods	Ensemble of MCDM and ML models for an aggregated Flood Susceptibility Index
[180]	MCDM +Deep Neural Networks; SHAP based XAI	National Flood Risk Insurance Data; Flood Risk Map. Improved risk at finer spatial levels
[181]	CART,MARS, BRT, SVM and linear Discriminant Analysis	Vulnerability using AHP (MCDM); Flood Hazards via ML. Different criteria used for vulnerability and hazards.

Multi-Criteria Decision Making (MCDM) approaches are used to aggregate such factors into a composite index. A suite of MCDM approaches such as the Simple Additive Weighting (SAW), Analytic Hierarchy Process (AHP), VIKOR, ELECTRE and others have been utilized for this purpose (see [182,183] for an overview of these techniques). Briefly, MCDM, aggregates ratings which describe an alternative's concordance with a criterion and criteria weights which denote the relative importance of the criterion and aggegates the scores over all selected criteria to rank and prioritize alternatives. Uncertainty is an integral component of such analysis and therefore some studies have attempted to use fuzzy set theory to capture differences among decision-makers with regards to different criteria and ratings of alternatives.

Note that MCDM is an unsupervised approach where in each alternative is ranked (so the number of initial clusters equals the number of alternatives. Once ranked, the grouping of these alternatives can be carried out subjectively or using unsupervised classification methods such as clustering and self-organizing maps (SOM).

While vulnerability captures the likely impact (or lack thereof) of floods at a given location. The hazard posed by the flood also varies geographically. This risk arises because of hydrologic conditions such as rainfall intensity and also land use land cover (LULC) conditions that lead to runoff (flood) generation. Risk of flooding of a land parcel depends upon several factors including but not limited to rainfall intensity, duration, LULC, soil types, antecedent moisture conditions and topographic slopes. Supervised machine learning models are used to develop relationships between flooded areas and other surrogate measures to model flooding hazard. A variety of approaches have been used to obtain flooded areas including but not limited to, use of satellite-imagery, household surveys, topography based wetness indices, and subjective or anecdotal information of decision makers. Clearly, the more direct the measurement, the more reliable is the result. A suite of machine learning modeling including but not limited to ANN, SVM, CART, MARS, Random Forest, Boosting Methods (GBT, XGBoost, AdaBoost) and deep learning have been employed for this purposes and demonstrated to model flood hazards with a high degree of accuracy.

The data volume available for modeling is often limited, which in turn favors conventional ML methods over deep learning paradigms. Some researchers have also utilized ensemble and stacking of algorithms with varying degree of success. The suitability of ensemble methods are hampered by the large predictive variability across different models. While most of the explanations of the ML modeling results is provided using tree-based approaches, the Shapely Additive Explanatins (SHAP) has also been used with deep learning models. While most models have focused on binary (Flood/No Flood) classification, multinomial models to capture varying levels of hazard have been explored to a limited extent.

Copula based approaches are widely used to estimate joint and conditional risks of flood intensity, volume and duration [184,185]. However, the integration of these approaches with machine learning is another poorly explored area of research. The coupling of ML and copula approaches to study droughts is demonstrated in [186], indicating the possibility of such an effort for flooding as well. Unlike droughts which depend upon atmospheric variables (rainfall and temperature), estimation of flood duration, volume and intensity is only possible in gaged catchments and the flood indicators also depend on soil type, antecedent moisture, land use and other factors whose temporal variability is not known in many locations.

Flood hazard assessment in many areas is hampered by paucity of data. This is particularly true not only in developing countries but also in rural and peri-urban areas of developed nations. Transfer learning provides one way to overcome this limitation where a model trained over a larger dataset in a similar region is used to make predictions in another. In addition, several other emerging technologies provide useful information to understand flooding hazards and vulnerability. The Internet of Things (IoT) is bringing about new, low-cost technologies for monitoring both fluvial and pluvial floods [187–189]. Low-cost sensors have the potential to significantly increase that data that are available for flood mapping. While such sensors also increase the data availability, they generally have poorer data

 doi:10.20944/preprints202504.1887.v1

quality and prone to giving erratic results. Therefore, data cleaning becomes necessary and machine learning algorithms for detecting and removing outliers can be useful [190]. In addition, several machine learning models both conventional and deep, can be used to make real-time forecasting and enhance the early flood warning capabilities of downstream communities. While the integration of IoT and machine learning or early warning systems has been demonstrated in some areas [191], there is still considerable opportunity to explore this area particularly in terms of data retrieval (using drones) when internet infrastructure may become non-functional due to floods.

7. Summary and Conclusions

Flooding is a complex problem that causes widespread damage across the world. It is now clear that flooding hazards cannot be completely eliminated using engineering measures alone. People have to live with floods. Flood control measures must not only focus on in-stream controls but also utilize structural and non-structural alternatives within an urban area to mitigate pluvial flooding to protect humans and also help mitigate downstream fluvial impacts. Despite keeping humans away from flood waters and managing them as necessary, the threats of flooding continue to grow unabated due to changes in climate, population growth and concomitant urbanization. Therefore, the current thinking goes a step further and evaluates how humans and the environment of a region are at risk of flooding and what are the best strategies to absorb the shocks of flooding and rebound to a better (or at least to the pre-flooding) state. The idea of withstanding the harmful effects of flooding to the best possible extent (robustness) and rebounding back to a better state as quickly as possible (rapidity) is called resiliency based flood management.

Resiliency-based flood management is not independent of flood control measures for minimizing fluvial and pluvial flooding, rather it takes a more holistic point of view to identify the risks that exist within a region and use it to guide future growth that is flood informed and also identify vulnerable targets that need immediate attention to bring the system back to normalcy. Thus, resiliency is not simply an engineering or land-planning and zoning endeavor but involves every one in the community to contribute towards flood-proofing.

Artificial intelligence (AI) in general and machine learning (ML) in particular can assist with resiliency based flood mitigation efforts. Machine learning models are universal approximators capable of capturing highly nonlinear phenomenon like floods. They overcome many deficiencies of physics-based models and are capable of providing better results with more easily available data in many cases. Machine learning is not a single technique but encompasses a wide range of algorithms that can be combined in more than one way to create solutions that are many times better than any individual algorithm that it is based on. These algorithms, especially newer deep learning based methods are quickly being adopted for flood studies. While discipline-specific of machine learning models have been reviewed in many articles on a regular pace, a holistic evaluation of machine learning in the context of flood studies is missing and a prime focus of this study.

The review not only restricts to evaluating machine learning models, but providing a context of flood management from command and control approaches to stakeholder-driven resiliency-risk based methods. The review also provides an overview of physics-based models that continue to be used today. These models often provide critical information such as inundation depth and extent that can be abstracted using machine learning schemes.

A review of literature indicates that many articles focus on a few ML techniques and identify a best predictor or classifier among them. As ML approaches are generally not taught in traditional hydrology classes, a review of various algorithms are presented to help understand their broad functioning. More in-depth reviews of the use of Machine learning for reservoir inflow and reservoir outflow and reservoir abstraction rules are presented. The LSTM model is widely used in both cases, but conventional methods are also in use. Levees and floodwalls are used to protect floodplains from inundation. Cracks in levees, piping and liquefaction are major causes of levee failure along with overtopping and compaction. CNN models and other object detection methods are being used to

 doi:10.20944/preprints202504.1887.v1

identify these zones of failures. Pump stations are operated to control flood waters from leaving riparian areas. Machine learning models are being explored to optimize the operation of these pump stations, the relatively small number of studies in this area directly corresponds to lack of suitable datasets.

A wide variety of machine learning models are being used to identify (unmapped) flooded areas based on a few known points (high water marks). In a similar vein, the results of the physics-based models are being abstracted by machine learning models to quickly compute flooding depths and the extent of inundation. Deep neural networks have shown considerable promise in this area. Machine learning models are also used to explain water quality transformations in a flood detention basins. Reinforcement learning in particular has been useful to control the releases from flood detention basins based on both quality and volume constraints. Flood resiliency planning studies often combine multi-criteria decision making (MCDM) modeling approaches to quantify vulnerability. Vulnerability is inversely related to resiliency and is a combination of disaster enhancing and disaster diminishing factors. In addition to MCDM, unsupervised learning approaches such as clustering have also been utilized. The flood hazard denotes the impact of the flood. Supervised learning methods are used to model flooding hazards. While many ML algorithms have been used, traditional approaches like SVM are seen to be better. The use of deep learning methods is constrained by the lack of availability of large volumes of data. In a similar vein, ensemble methods have also had limited success because of extreme variability in predictions.

The review also identified several unexplored or poorly explored directions in flood resiliency studies. In particular, capturing large-scale uncertainty in reservoir inflow predictions due to climate change is one such area. The use of ML for assessing risks to levees and floodwalls by integrating drone-based survey with computer vision and machine learning is another area. Similarly, the paucity of models for modeling pump station operations is limited by lack of suitable datasets. The development of such a dataset and providing in the public-domain will improve the number of applications in this area. In a similar vein, identification of pluvial flooding locations often relies on the use of physics-based model for quantifying the extent and nature of inundation. Therefore, the accuracy of ML models using this data cannot be higher than the physical-models they are based on. Remote-sensing methods particularly using Interferometric Synthetic Aperture Radar (InSAR) can be valuable along with improved field data collection to better capture inundation risks.

The use of deep learning and ensemble based methods for flood hazard identification has also seen limited success in flood risk assessment studies, especially for capturing fluvial flood risks. This limitation largely occurs due to paucity of data. Deep learning algorithms require considerable amounts of data and when such data are available (e.g., reservoir outflow in a stream) their predictive superiority becomes evident. Coupling of ML models with other statistical risk assessment methods such as copula theory has also been hindered by the lack of suitable data and presents another avenue to explore to move flood risk from a largely binary (flood/no flood) univariate calculation to its assessment over multiple dimensions.

In closing, machine learning models are widely being used in flood studies ranging from flood control to resiliency-based management. The usage is higher in those applications where data are readily available. Advances in remote-sensing, drone based data collection and proliferation of low-cost sensors using the Internet-of-Things (IoT) will further expand its role in many other aspects of floods that are currently hampered by data limitations. While machine learning offers great prediction and forecasting abilities, there is no single universally acceptable method, so empirical experimentation with multiple models becomes necessary. This aspect should not be viewed as a weakness but exploited to learn more about the functioning of the models and what is making them better predictors (or not). Explainable AI (XAI) methods come in handy in this regard and should be adopted to transition ML models for black-box forecasters to better informed and knowledge-based predictors.

Author Contributions: Conceptualization, V.U. and EA.H.; methodology, EA.H., V.U.; formal analysis, VU,EAH; investigation, EAH,VU; resources, VU,EAH.; data curation, EA.H; writing—original draft preparation, V.U.;

 doi:10.20944/preprints202504.1887.v1

writing—review and editing, EA.H.; visualization, EA.H.,V.U.; supervision, EA.H.; project administration, V.U. All authors have read and agreed to the published version of the manuscript.

Funding: None.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Kreibich, H.; Dimitrova, B. Assessment of damages caused by different flood types. *WIT Transactions on Ecology and the Environment* **2010**, *133*, 3–11.
2. USDHS2025. United States Department of Homeland Security - Natural Disasters, 2025.
3. Ruan, X.; Sun, H.; Shou, W.; Wang, J. The impact of climate change and urbanization on compound flood risks in coastal areas: a comprehensive review of methods. *Applied Sciences* **2024**, *14*, 10019.
4. Tabari, H. Climate change impact on flood and extreme precipitation increases with water availability. *Scientific reports* **2020**, *10*, 13768.
5. Saghafian, B.; Farazjoo, H.; Bozorgy, B.; Yazdandoost, F. Flood intensification due to changes in land use. *Water resources management* **2008**, *22*, 1051–1067.
6. Maranzoni, A.; D'Oria, M.; Rizzo, C. Probabilistic mapping of life loss due to dam-break flooding. *Natural Hazards* **2024**, *120*, 2433–2460.
7. Montalvo, C.; Reyes-Silva, J.; Sañudo, E.; Cea, L.; Puertas, J. Urban pluvial flood modelling in the absence of sewer drainage network data: A physics-based approach. *Journal of Hydrology* **2024**, *634*, 131043.
8. Okoli, K.; Mazzoleni, M.; Breinl, K.; Di Baldassarre, G. A systematic comparison of statistical and hydrological methods for design flood estimation. *Hydrology Research* **2019**, *50*, 1665–1678.
9. Yan, B.; Mu, R.; Guo, J.; Liu, Y.; Tang, J.; Wang, H. Flood risk analysis of reservoirs based on full-series ARIMA model under climate change. *Journal of Hydrology* **2022**, *610*, 127979.
10. Latif, S.; Mustafa, F. Copula-based multivariate flood probability construction: a review. *Arabian Journal of Geosciences* **2020**, *13*, 132.
11. Wong, W.; Lee, M.; Azman, A.; Rose, L.; Teknousahawan, F. Development of Short-term Flood Forecast Using ARIMA. *International Journal of Mathematical Models and Methods in Applied Sciences* **2021**, *15*, 68–75.
12. Soares, J.A.; Ozelim, L.C.; Bacelar, L.; Ribeiro, D.B.; Stephany, S.; Santos, L.B. ML4FF: A machine-learning framework for flash flood forecasting applied to a Brazilian watershed. *Journal of Hydrology* **2025**, *652*, 132674.
13. Inc., A. Google Scholar, 2025.
14. Ghorpade, P.; Gadge, A.; Lende, A.; Chordiya, H.; Gosavi, G.; Mishra, A.; Hooli, B.; Ingle, Y.; Shaikh, N. Flood forecasting using machine learning: a review. In Proceedings of the 2021 8th international conference on smart computing and communications (ICSCC. IEEE, 2021, pp. 32–36.
15. Mosavi, A.; Ozturk, P.; Chau, K. Flood prediction using machine learning models: Literature review. *Water* **2018**, *10*, 1536.
16. Ferrari, R. Writing narrative style literature reviews. *Medical Research* **2015**, *24*, 230–234.
17. Pollock, A.; Berge, E. How to do a Systematic Review. *Internationl Journal of Stroke* **2018**, *13*, 138–156.
18. Kundzewicz, Z.; Su, B.; Wang, Y.; Xia, J.; Huang, J.; Jiang, T. Flood risk and its reduction in China. *Advances in Water Resources* **2019**, *130*, 37–45.
19. Ramirez, J.; Adamowicz, W.; Easter, K.; Graham-Tomasi, T. Ex post analysis of flood control: Benefit-cost analysis and the value of information. *Water Resources Research* **1988**, *24*, 1397–1405.
20. García-Ledesma, I.; Madrigal, J.; Domínguez-Sánchez, C.; Sánchez-Quispe, S.; Lara-Ledesma, B. Importance of cost-benefit evaluation in the selection of flood control infrastructures. *Urban Water Journal* **2025**, pp. 1–13.
21. Nguyen, B.; Binh, D.; Tran, T.; Kantoush, S.; Sumi, T. Response of streamflow and sediment variability to cascade dam development and climate change in the Sai Gon Dong Nai River basin. *Climate Dynamics* **2024**, *62*, 7997–8017.
22. Eckart, K.; McPhee, Z.; Bolisetti, T. Performance and implementation of low impact development–A review. *Science of the Total Environment* **2017**, *607*, 413–432.
23. Liu, T.; Lawluvy, Y.; Shi, Y.; Yap, P. Low impact development (LID) practices: A review on recent developments, challenges and prospects. *Water, Air, & Soil Pollution* **2021**, *232*, 344.
24. Nasiri Khiavi, A.; Vafakhah, M.; Sadeghi, S.; Jun, C.; Bateni, S. Comparative effect of traditional and collaborative watershed management approaches on flood components. *Journal of Flood Risk Management* **2025**, *18*, 13037.

25. Grigg, N. Two Decades of Integrated Flood Management: Status, Barriers, and Strategies. *Climate* **2024**, *12*, 67.
26. Thampapillai, D.; Musgrave, W. Flood damage mitigation: A review of structural and nonstructural measures and alternative decision frameworks. *Water Resources Research* **1985**, *21*, 411–424.
27. Kaya, C.; Derin, L. Parameters and methods used in flood susceptibility mapping: a review. *Journal of Water and Climate Change* **2023**, *14*, 1935–1960.
28. Reimann, L.; Vafeidis, A.; Honsel, L. Population development as a driver of coastal risk: Current trends and future pathways. *Cambridge Prisms: Coastal Futures* **2023**, *1*, 14.
29. Kammu, M.; De Moel, H.; Salvucci, G.; Viviroli, D.; Ward, P.J.; Varis, O. Over the Hills and further away from the coast: Global geospatial patterns of human and the environment over the 20th - 21st centuries. *Environmental Research Letters* **2016**, *11*, 0304010.
30. Kron, W. Flood risk= hazard• values• vulnerability. *Water international* **2005**, *30*, 58–68.
31. Peters, D.; Caissie, D.; Monk, W.; Rood, S.; St-Hilaire, A. An ecological perspective on floods in Canada. *Canadian Water Resources Journal/Revue canadienne des ressources hydriques* **2016**, *41*, 288–306.
32. Parsons, M. Extreme floods and river values: A social–ecological perspective. *River Research and Applications* **2019**, *35*, 1677–1687.
33. Owusu, A.; Mul, M.; Zaag, P.; Slinger, J. Re-operating dams for environmental flows: From recommendation to practice. *River Research and Applications* **2021**, *37*, 176–186.
34. Graha, D.; Yudono, A.; Afrianto, F. IHST (Integrated, Holistic, Spatial, and Thematic) Flood Management Model: The Integration of Flood Model, Green Infrastructure and Non-Structural Mitigation in the Urban Area of Barabai City. *IOP Conference Series: Earth and Environmental Science* **2024**, *1391*, 012026. Publisher: IOP Publishing.
35. Laidlaw, S.; Percival, S. Flood resilience: a review of evolving definitions. *Natural Hazards* **2024**, *120*, 10773–10784.
36. Martin-Breen, P.; Anderies, J. Resilience: a literature review. Report –, Institute of Development Studies, 2011.
37. McClymont, K.; Morrison, D.; Beevers, L.; Carmen, E. Flood resilience: a systematic review. *Journal of Environmental Planning and Management* **2020**, *63*, 1151–1176.
38. Wang, L.; Cui, S.; Li, Y.; Huang, H.; Manandhar, B.; Nitivattananon, V.; Fang, X.; Huang, W. A review of the flood management: from flood control to flood resilience. *Heliyon* **2022**, *8*.
39. Kafle, A.; Hernandez, E.A.; Uddameri, V. Resiliency of Hydraulic Infrastructure Designs in a Climate Hot-Spot at the Intersection of Two Climate Zones. *Natural Hazard Research* **2025**, *In-Press*. https://doi.org/https://doi.org/10.1016/j.nhres.2025.03.010.
40. Shi, L.; Fisher, A.; Brenner, R.; Greiner-Safi, A.; Shepard, C.; Vanucchi, J. Equitable buyouts? Learning from state, county, and local floodplain management programs. *Climatic Change* **2022**, *174*, 29.
41. Liu, S.; Huang, S.; Xie, Y.; Wang, H.; Leng, G.; Huang, Q.; Wei, X.; Wang, L. Identification of the non-stationarity of floods: changing patterns, causes, and implications. *Water Resources Management* **2019**, *33*, 939–953.
42. Peel, M.C.; McMahon, T.A. Historical development of rainfall-runoff modeling. *Wiley Interdisciplinary Reviews: Water* **2020**, *7*, e1471.
43. Ning, L.; Zhan, C.; Luo, Y.; Wang, Y.; Liu, L. A review of fully coupled atmpshere-hydrology simulations. *Journal of Geographical Sciences* **2019**, *29*, 465–479.
44. Wada, Y.; Bierkens, M.F.; De Roo, A.; Dirmeyer, P.A.; Famiglietti, J.S.; Hanasaki, N.; Konar, M.; Liu, J.; Müller Schmied, H.; Oki, T.; et al. Human–water interface in hydrological modelling: current status and future directions. *Hydrology and Earth System Sciences* **2017**, *21*, 4169–4193. Publisher: Copernicus GmbH.
45. Miralles-Wilhelm, F. Development and application of integrative modeling tools in support of food-energy-water nexus planning—a research agenda. *Journal of Environmental Studies and Sciences* **2016**, *6*, 3–10. Publisher: Springer.
46. Oreskes, N.; Shrader-Frechette, K.; Belitz, K. Verification, validation, and confirmation of numerical models in the earth sciences. *Science* **1994**, *263*, 641–646.
47. Jakeman, A.; Hornberger, G. How much complexity is warranted in a rainfall-runoff model? *Water resources research* **1993**, *29*, 2637–2649.
48. Song, J.; Her, Y.; Kang, M. Estimating reservoir inflow and outflow from water level observations using expert knowledge: dealing with an ill-posed water balance equation in reservoir management. *Water Resources Research* **2022**, *58*, 2020 028183.

49. de Lavenne, A.; Andréassian, V.; Thirel, G.; Ramos, M.; Perrin, C. A regularization approach to improve the sequential calibration of a semidistributed hydrological model. *Water Resources Research* **2019**, *55*, 8821–8839.
50. Sun, Y.; Bao, W.; Jiang, P.; Si, W.; Zhou, J.; Zhang, Q. Development of a regularized dynamic system response curve for real-time flood forecasting correction. *Water* **2018**, *10*, 450.
51. Zoccatelli, D.; Wright, D.; White, J.; Fienen, M.; Yu, G. Precipitation uncertainty estimation and rainfall-runoff model calibration using iterative ensemble smoothers. *Advances in Water Resources* **2024**, *186*, 104658.
52. Asgari, M.; Yang, W.; Lindsay, J.; Tolson, B.; Dehnavi, M.M. A review of parallel computing applications in calibrating watershed hydrologic models. *Environmental Modelling & Software* **2022**, *151*, 105370.
53. Teng, J.; Jakeman, A.; Vaze, J.; Croke, B.; Dutta, D.; Kim, S. Flood inundation modelling: A review of methods, recent advances and uncertainty analysis. *Environmental modelling & software* **2017**, *90*, 201–216.
54. Sarchani, S.; Seiradakis, K.; Coulibaly, P.; Tsanis, I. Flood inundation mapping in an ungauged basin. *Water* **2020**, *12*, 1532.
55. Guo, Y.; Zhang, Y.; Zhang, L.; Wang, Z. Regionalization of hydrological modeling for predicting streamflow in ungauged catchments: A comprehensive review. *Wiley Interdisciplinary Reviews: Water* **2021**, *8*, 1487.
56. Pool, S.; Vis, M.; Seibert, J. Regionalization for ungauged catchments—lessons learned from a comparative large-sample study. *Water Resources Research* **2021**, *57*, 2021 030437.
57. Blöschl, G.; Sivapalan, M. Scale issues in hydrological modelling: a review. *Hydrological processes* **1995**, *9*, 251–290.
58. Neri, M.; Parajka, J.; Toth, E. Importance of the informative content in the study area when regionalising rainfall-runoff model parameters: the role of nested catchments and gauging station density. *Hydrology and Earth System Sciences* **2020**, *24*, 5149–5171. Publisher: Copernicus Publications Göttingen, Germany.
59. Cerbelaud, A.; David, C.H.; Biancamaria, S.; Wade, J.; Tom, M.; Prata de Moraes Frasson, R.; Blumstein, D. Peak flow event durations in the Mississippi River basin and implications for temporal sampling of rivers. *Geophysical Research Letters* **2024**, *51*, e2024GL109220.
60. Neal, J.; Villanueva, I.; Wright, N.; Willis, T.; Fewtrell, T.; Bates, P. How much physical complexity is needed to model flood inundation? *Hydrological Processes* **2012**, *26*, 2264–2282.
61. Gupta, H.; Kling, H.; Yilmaz, K.; Martinez, G. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of hydrology* **2009**, *377*, 80–91.
62. Mizukami, N.; Rakovec, O.; Newman, A.; Clark, M.; Wood, A.; Gupta, H.; Kumar, R. On the choice of calibration metrics for "high-flow" estimation using hydrologic models. *Hydrology and Earth System Sciences* **2019**, *23*, 2601–2614.
63. Bárdossy, A.; Anwar, F. Why our rainfall-runoff models keep underestimating the peak flows? *Hydrology and Earth System Sciences Discussions* **2022**, pp. 1–30.
64. Samuel, A. Some studies in machine learning using the game of checkers. *IBM Journal of research and development* **1959**, *3*, 210–229.
65. Liu, S.; Liu, R.; Tan, N. A spatial improved-kNN-based flood inundation risk framework for urban tourism under two rainfall scenarios. *Sustainability* **2021**, *13*, 2859.
66. Crespo, J.; Mora, E. Drought estimation with neural networks. *Advances in Engineering Software* **1993**, *18*, 167–170.
67. Karunanithi, N.; Grenney, W.; Whitley, D.; Bovee, K. Neural networks for river flow prediction. *Journal of computing in civil engineering* **1994**, *8*, 201–220.
68. Hughes, J.; Lettenmaier, D.; Wood, E. An approach for assessing the sensitivity of floods to regional climate change. *AIP Conference Proceedings-CONF:9201138* **1992**, *277*, 112–124.
69. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research* **2011**, *12*, 2825–2830. Publisher: JMLR. org.
70. Pang, B.; Nijkamp, E.; Wu, Y.N. Deep learning with tensorflow: A review. *Journal of Educational and Behavioral Statistics* **2020**, *45*, 227–248. Publisher: SAGE Publications Sage CA: Los Angeles, CA.
71. Kaelbling, L.; Littman, M.; Moore, A. Reinforcement learning: A survey. *Journal of artificial intelligence research* **1996**, *4*, 237–285.
72. Paraskevopoulos, E.; Anagnostopoulou, A.; Chalas, N.; Karagianni, M.; Bamidis, P. Unravelling the multisensory learning advantage: Different patterns of within and across frequency-specific interactions drive uni-and multisensory neuroplasticity. *NeuroImage* **2024**, *291*, 120582.
73. Robbins, H.; Monro, S. A stochastic approximation method. *The annals of mathematical statistics* **1951**, pp. 400–407.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 April 2025 doi:10.20944/preprints202504.1887.v1

74. Gebremedhin, A.H.; Walther, A. An introduction to algorithmic differentiation. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **2020**, *10*, e1334.
75. Dasgupta, S. Analysis of a greedy active learning strategy. *Advances in neural information processing systems* **2004**, *17*.
76. Parimbelli, E.; Buonocore, T.M.; Nicora, G.; Michalowski, W.; Wilk, S.; Bellazzi, R. Why did AI get this one wrong?—Tree-based explanations of machine learning model predictions. *Artificial intelligence in medicine* **2023**, *135*, 102471.
77. Belle, V.; Papantonis, I. Principles and practice of explainable machine learning. *Frontiers in big Data* **2021**, *4*, 688969.
78. García-Feal, O.; González-Cao, J.; Fernández-Nóvoa, D.; Astray Dopazo, G.; Gómez-Gesteira, M. Comparison of machine learning techniques for reservoir outflow forecasting. *Natural Hazards and Earth System Sciences Discussions* **2022**, pp. 1–27.
79. Castillo-Botón, C.; Casillas-Pérez, D.; Casanova-Mateo, C.; Moreno-Saavedra, L.; Morales-Díaz, B.; Sanz-Justo, J.; Gutiérrez, P.; Salcedo-Sanz, S. Analysis and prediction of dammed water level in a hydropower reservoir using machine learning and persistence-based techniques. *Water* **2020**, *12*, 1528.
80. Latif, S.; Ahmed, A.; Sherif, M.; Sefelnasr, A.; El-Shafie, A. Reservoir water balance simulation model utilizing machine learning algorithm. *Alexandria Engineering Journal* **2021**, *60*, 1365–1378.
81. Liu, Y.; Qin, H.; Zhang, Z.; Yao, L.; Wang, Y.; Li, J.; Liu, G.; Zhou, J. Deriving reservoir operation rule based on Bayesian deep learning method considering multiple uncertainties. *Journal of Hydrology* **2019**, *579*, 124207.
82. Soria-Lopez, A.; Sobrido-Pouso, C.; Mejuto, J.; Astray, G. Assessment of different machine learning methods for reservoir outflow forecasting. *Water* **2023**, *15*, 3380.
83. Qie, G.; Zhang, Z.; Getahun, E.; Allen Mamer, E. Comparison of machine learning models performance on simulating reservoir outflow: A case study of two reservoirs in Illinois, USA. *JAWRA Journal of the American Water Resources Association* **2023**, *59*, 554–570.
84. Al-Nouti, A.F.; Fu, M.; Bokde, N.D. Reservoir operation based machine learning models: comprehensive review for limitations, research gap, and possible future research direction. *Knowledge-Based Engineering and Sciences* **2024**, *5*, 75–139.
85. Chen, R.; Wang, D.; Mei, Y.; Lin, Y.; Lin, Z.; Zhang, Z.; Zhuang, S. A knowledge-guided LSTM reservoir outflow model and its application to streamflow simulation in reservoir-regulated basins. *Journal of Hydrology* **2025**, p. 133164.
86. Yi, S.; Yi, J. Reservoir-based flood forecasting and warning: deep learning versus machine learning. *Applied Water Science* **2024**, *14*, 1–23.
87. Tilloy, A.; Paprotny, D.; Grimaldi, S.; Gomes, G.; Bianchi, A.; Lange, S.; Beck, H.; Mazzetti, C.; Feyen, L. HERA: a high-resolution pan-European hydrological reanalysis (1951–2020. *Earth System Science Data* **2025**, *17*, 293–316.
88. Yang, S.; Yang, D.; Chen, J.; Zhao, B. Real-time reservoir operation using recurrent neural networks and inflow forecast from a distributed hydrological model. *Journal of Hydrology* **2019**, *579*, 124229.
89. Hoedt, P.; Kratzert, F.; Klotz, D.; Halmich, C.; Holzleitner, M.; Nearing, G.; Hochreiter, S.; Klambauer, G. Mc-lstm: Mass-conserving lstm. In Proceedings of the International conference on machine learning PMLR, 2021, pp. 4275–4286.
90. Pokharel, S.; Roy, T.; Admiraal, D. Effects of mass balance, energy balance, and storage-discharge constraints on LSTM for streamflow prediction. *Environmental modelling & software* **2023**, *166*, 105730.
91. Fan, M.; Zhang, L.; Liu, S.; Yang, T.; Lu, D. Investigation of hydrometeorological influences on reservoir releases using explainable machine learning methods. *Frontiers in Water* **2023**, *5*, 1112970.
92. Fan, M.; Liu, S.; Lu, D.; Gangrade, S.; Kao, S. Explainable machine learning model for multi-step forecasting of reservoir inflow with uncertainty quantification. *Environmental Modelling & Software* **2023**, *170*, 105849.
93. Sushanth, K.; Mishra, A.; Mukhopadhyay, P.; Singh, R. Real-time streamflow forecasting in a reservoir-regulated river basin using explainable machine learning and conceptual reservoir module. *Science of the Total Environment* **2023**, *861*, 160680.
94. Rajesh, M.; Anishka, S.; Viksit, P.; Arohi, S.; Rehana, S. Improving short-range reservoir inflow forecasts with machine learning model combination. *Water Resources Management* **2023**, *37*, 75–90.
95. Huang, I.; Chang, M.; Lin, G. An optimal integration of multiple machine learning techniques to real-time reservoir inflow forecasting. *Stochastic Environmental Research and Risk Assessment* **2022**, *36*, 1541–1561.

96. Tian, D.; He, X.; Srivastava, P.; Kalin, L. A hybrid framework for forecasting monthly reservoir inflow based on machine learning techniques with dynamic climate forecasts, satellite-based data, and climate phenomenon information. *Stochastic Environmental Research and Risk Assessment* **2021**, pp. 1–23.
97. Zhang, W.; Wang, H.; Lin, Y.; Jin, J.; Liu, W.; An, X. Reservoir inflow predicting model based on machine learning algorithm via multi-model fusion: A case study of Jinshuitan river basin. *IET Cyber-Systems and Robotics* **2021**, *3*, 265–277.
98. Latif, S.; Ahmed, A. A review of deep learning and machine learning techniques for hydrological inflow forecasting. *Environment, Development and Sustainability* **2023**, *25*, 12189–12216.
99. Gupta, A.; Kumar, A. Two-step daily reservoir inflow prediction using ARIMA-machine learning and ensemble models. *Journal of Hydro-environment Research* **2022**, *45*, 39–52.
100. Ibrahim, K.; Huang, Y.; Ahmed, A.; Koo, C.; El-Shafie, A. Forecasting multi-step-ahead reservoir monthly and daily inflow using machine learning models based on different scenarios. *Applied Intelligence* **2023**, *53*, 10893–10916.
101. Paul, T.; Raghavendra, S.; Ueno, K.; Ni, F.; Shin, H.; Nishino, K.; Shingaki, R. Forecasting of reservoir inflow by the combination of deep learning and conventional machine learning. In Proceedings of the 2021 international conference on data mining workshops (ICDMW. IEEE, 2021, pp. 558–565.
102. Deb, D.; Arunachalam, V.; Raju, K. Daily reservoir inflow prediction using stacking ensemble of machine learning algorithms. *Journal of Hydroinformatics* **2024**, *26*, 972–997.
103. Fan, M.; Liu, S.; Lu, D. Advancing subseasonal reservoir inflow forecasts using an explainable machine learning method. *Journal of Hydrology: Regional Studies* **2023**, *50*, 101584.
104. Luo, B.; Fang, Y.; Wang, H.; Zang, D. Reservoir inflow prediction using a hybrid model based on deep learning. *IOP Conference Series: Materials Science and Engineering* **2020**, *715*, 012044. Publisher: IOP Publishing.
105. Ahmadi, F.; Ghasemlounia, R.; Gharehbaghi, A. Machine learning approaches coupled with variational mode decomposition: a novel method for forecasting monthly reservoir inflows. *Earth Science Informatics* **2024**, *17*, 745–760.
106. Vasheghani Farahani, E.; Massah Bavani, A.; Roozbahani, A. Enhancing reservoir inflow forecasting precision through Bayesian Neural Network modeling and atmospheric teleconnection pattern analysis. *Stochastic Environmental Research and Risk Assessment* **2025**, *39*, 205–229.
107. Noorbeh, P.; Roozbahani, A.; Kardan Moghaddam, H. Annual and monthly dam inflow prediction using Bayesian networks. *Water Resources Management* **2020**, *34*, 2933–2951.
108. Elzain, H.; Abdalla, O.; Al-Maktoumi, A.; Kacimov, A.; Eltayeb, M. A novel approach to forecast water table rise in arid regions using stacked ensemble machine learning and deep artificial intelligence models. *Journal of Hydrology* **2024**, *640*, 131668.
109. Liu, Y.; Qin, H.; Zhang, Z.; Yao, L.; Wang, Y.; Li, J.; Liu, G.; Zhou, J. Deriving reservoir operation rule based on Bayesian deep learning method considering multiple uncertainties. *Journal of Hydrology* **2019**, *579*, 124207.
110. Zhang, X.; Wang, H.; Peng, A.; Wang, W.; Li, B.; Huang, X. Quantifying the uncertainties in data-driven models for reservoir inflow prediction. *Water Resources Management* **2020**, *34*, 1479–1493.
111. Flynn, S.; Zamanian, S.; Vahedifard, F.; Shafieezadeh, A.; Schaaf, D. Data-Driven Model for Estimating the Probability of Riverine Levee Breach Due to Overtopping. *Journal of Geotechnical and Geoenvironmental Engineering* **2022**, *148*, 04021193.
112. Kuchi, A.; Panta, M.; Hoque, M.; Abdelguerfi, M.; Flanagin, M. A machine learning approach to detecting cracks in levees and floodwalls. *Remote Sensing Applications: Society and Environment* **2021**, *22*, 100513.
113. Zhao, X.; Zhang, H.; Wang, P.; Ren, Q.; Zhang, D. Improving efficiency and accuracy of levee hazard detection with deep learning. *Computers & Geosciences* **2024**, *187*, 105593.
114. Russo, B.; Athanasopoulos-Zekkos, A. Exploration of feature engineering techniques and unsupervised machine learning clustering algorithms for geophysical data on levees. In *Geo-Congress 2024*; 2024; pp. 454–463.
115. Alzubaidi, L.; Chlaib, H.; Fadhel, M.; Chen, Y.; Bai, J.; Albahri, A.; Gu, Y. Reliable deep learning framework for the ground penetrating radar data to locate the horizontal variation in levee soil compaction. *Engineering Applications of Artificial Intelligence* **2024**, *129*, 107627.
116. Kuchi, A.; Hoque, M.; Abdelguerfi, M.; Flanagin, M. Machine learning applications in detecting sand boils from images. *Array* **2019**, *3*, 100012.
117. U.S. Army Corps of Engineers (USACE). National Levee Database. https://levees.sec.usace.army.mil/about/about-the-data/, 2025. Accessed: April 15, 2025.

 doi:10.20944/preprints202504.1887.v1

118. Lee, W.; Lee, E. Runoff prediction based on the discharge of pump stations in an urban stream using a modified multi-layer perceptron combined with meta-heuristic optimization. *Water* **2022**, *14*, 99.
119. Wang, W.; Sang, G.; Zhao, Q.; Lu, L. Water level prediction of pumping station pre-station based on machine learning methods. *Water Supply* **2023**, *23*, 4092–4111.
120. Kow, P.Y.; Liou, J.Y.; Yang, M.T.; Lee, M.H.; Chang, L.C.; Chang, F.J. Advancing climate-resilient flood mitigation: Utilizing transformer-LSTM for water level forecasting at pumping stations. *Science of the Total Environment* **2024**, *927*, 172246.
121. Joo, J.; Jeong, I.; Kang, S. Deep Reinforcement Learning for Multi-Objective Real-Time Pump Operation in Rainwater Pumping Stations. *Water* **2024**, *16*, 3398.
122. Choo, Y.; Kim, J.; Park, S.; Choo, T.; Choe, Y. Method for operating drainage pump stations considering downstream water level and reduction in urban river flooding. *Water* **2021**, *13*, 2741.
123. Broome, M.A. General Principles of Pumping Station Design and Layout. Technical Report CECW-EE Engineer Manual 1110-2-3102, U. S. Army Corps of Engineers, Washington, DC, 1995.
124. Aderyani, F.; Jafarzadegan, K.; Moradkhani, H. A surrogate machine learning modeling approach for enhancing the efficiency of urban flood modeling at metropolitan scales. *Sustainable Cities and Society* **2025**, *123*, 106277.
125. Tetteh, A.; Moomen, A.; Yevugah, L.; Tengnibuor, A. Geospatial approach to pluvial flood-risk and vulnerability assessment in Sunyani Municipality. *Heliyon* **2024**, *10*.
126. McSpadden, D.; Goldenberg, S.; Roy, B.; Schram, M.; Goodall, J.; Richter, H. A comparison of machine learning surrogate models of street-scale flooding in Norfolk, Virginia. *Machine Learning with Applications* **2024**, *15*, 100518.
127. Adeke, D.; Mugume, S. A methodology for development of flood-depth-velocity damage functions for improved estimation of pluvial flood risk in cities. *Journal of Hydrology* **2025**, *p.132736*.
128. Chen, G.; Hou, J.; Liu, Y.; Xue, S.; Wu, H.; Wang, T.; Lv, J.; Jing, J.; Yang, S. Urban inundation rapid prediction method based on multi-machine learning algorithm and rain pattern analysis. *Journal of Hydrology* **2024**, *633*, 131059.
129. Mehedi, M.A.A.; Smith, V.; Hosseiny, H.; Jiao, X. Unraveling the complexities of urban fluvial flood hydraulics through AI. *Scientific Reports* **2022**, *12*, 18738.
130. Rasool, U.; Yin, X.; Xu, Z.; Padulano, R.; Rasool, M.; Siddique, M.; Hassan, M.; Senapathi, V. Rainfall-driven machine learning models for accurate flood inundation mapping in Karachi, Pakistan. *Urban Climate* **2023**, *49*, 101573.
131. Bourget, M.; Boudreault, M.; Carozza, D.; Boudreault, J.; Raymond, S. A data science approach to climate change risk assessment applied to pluvial flood occurrences for the United States and Canada. *ASTIN Bulletin: The Journal of the IAA* **2024**, *54*, 495–517.
132. Ye, C.; Xu, Z.; Liao, W.; Li, X.; Shu, X. Capturing Urban Pluvial River Flooding Features Based on the Fusion of Physically Based and Data-Driven Approaches. *Sustainability* **2025**, *17*, 2524.
133. Katti, A.; Ashish, K.; Loke, A.; Bade, K. A Pluvial Flood Detection Model Using Machine Learning Techniques and Simulate The Flow of Water. In Proceedings of the 2020 5th International Conference on Communication and Electronics Systems (ICCES. IEEE, 2020, pp. 1189–1195.
134. Gao, W.; Liao, Y.; Chen, Y.; Lai, C.; He, S.; Wang, Z. Enhancing transparency in data-driven urban pluvial flood prediction using an explainable CNN model. *Journal of Hydrology* **2024**, *645*, 132228.
135. Burrichter, B.; Hofmann, J.; Silva, J.; Niemann, A.; Quirmbach, M. A spatiotemporal deep learning approach for urban pluvial flood forecasting with multi-source data. *Water* **2023**, *15*, 1760.
136. Allegri, E.; Zanetti, M.; Torresan, S.; Critto, A. Pluvial flood risk assessment for 2021–2050 under climate change scenarios in the Metropolitan City of Venice. *Science of the Total Environment* **2024**, *914*, 169925.
137. Safaei-Moghadam, A.; Hosseinzadeh, A.; Minsker, B. Predicting real-time roadway pluvial flood risk: A hybrid machine learning approach coupling a graph-based flood spreading model, historical vulnerabilities, and Waze data. *Journal of Hydrology* **2024**, *637*, 131406.
138. Hassan, T.; Majeed, S.; Memon, M. Urban Pluvial Flood Prediction Using Machine Learning Models. In Proceedings of the 2024 4th International Conference on Innovations in Computer Science (ICONICS. IEEE, 2024, pp. 1–6.
139. Liao, Y.; Wang, Z.; Chen, X.; Lai, C. Fast simulation and prediction of urban pluvial floods using a deep convolutional neural network model. *Journal of Hydrology* **2023**, *624*, 129945. Publisher: Elsevier.
140. Hofmann, J.; Schüttrumpf, H. Floodgan: Using deep adversarial learning to predict pluvial flooding in real time. *Water* **2021**, *13*, 2255. Publisher: Multidisciplinary Digital Publishing Institute.

141. Lin, J.; Zhang, W.; Wen, Y.; Qiu, S. Evaluating the association between morphological characteristics of urban land and pluvial floods using machine learning methods. *Sustainable Cities and Society* **2023**, *99*, 104891.
142. Zahura, F.; Goodall, J. Predicting combined tidal and pluvial flood inundation using a machine learning surrogate model. *Journal of Hydrology: Regional Studies* **2022**, *41*, 101087.
143. Ke, Q.; Tian, X.; Bricker, J.; Tian, Z.; Guan, G.; Cai, H.; Huang, X.; Yang, H.; Liu, J. Urban pluvial flooding prediction by machine learning approaches–a case study of Shenzhen city, China. *Advances in Water Resources* **2020**, *145*, 103719.
144. Hou, J.; Zhou, N.; Chen, G.; Huang, M.; Bai, G. Rapid forecasting of urban flood inundation using multiple machine learning models. *Natural Hazards* **2021**, *108*, 2335–2356.
145. Yan, X.; Xu, K.; Feng, W.; Chen, J. A rapid prediction model of urban flood inundation in a high-risk area coupling machine learning and numerical simulation approaches. *International Journal of Disaster Risk Science* **2021**, *12*, 903–918.
146. Rangari, V.; Umamahesh, N.; Bhatt, C. Assessment of inundation risk in urban floods using HEC RAS 2D. *Modeling Earth Systems and Environment* **2019**, *5*, 1839–1851.
147. Konapala, G.; Kumar, S.; Ahmad, S. Exploring Sentinel-1 and Sentinel-2 diversity for flood inundation mapping using deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing* **2021**, *180*, 163–173.
148. Afzal, M.; Ali, S.; Nazeer, A.; Khan, M.; Waqas, M.; Aslam, R.; Cheema, M.; Nadeem, M.; Saddique, N.; Muzammil, M.; et al. Flood inundation modeling by integrating HEC–RAS and satellite imagery: a case study of the Indus River Basin. *Water* **2022**, *14*, 2984.
149. Annis, A.; Nardi, F. GFPLAIN and multi-source data assimilation modeling: conceptualization of a flood forecasting framework supported by hydrogeomorphic floodplain rapid mapping. *Hydrology* **2021**, *8*, 143.
150. Wei, J.; Luo, X.; Huang, H.; Liao, W.; Lei, X.; Zhao, J.; Wang, H. Enable high-resolution, real-time ensemble simulation and data assimilation of flood inundation using distributed GPU parallelization. *Journal of Hydrology* **2023**, *619*, 129277.
151. Seleem, O.; Ayzel, G.; Bronstert, A.; Heistermann, M. Transferability of data-driven models to predict urban pluvial flood water depth in Berlin, Germany. *Natural Hazards and Earth System Sciences Discussions* **2022**, pp. 1–23.
152. Yang, Y.; Li, Y.; Huang, Q.; Xia, J.; Li, J. Surrogate-based multiobjective optimization to rapidly size low impact development practices for outflow capture. *Journal of Hydrology* **2023**, *616*, 128848.
153. Mullapudi, A.; Lewis, M.; Gruden, C.; Kerkez, B. Deep reinforcement learning for the real time control of stormwater systems. *Advances in water resources* **2020**, *140*, 103600.
154. Saliba, S.; Bowes, B.; Adams, S.; Beling, P.; Goodall, J. Deep reinforcement learning with uncertain data for real-time stormwater system control and flood mitigation. *Water* **2020**, *12*, 3222.
155. Bowes, B.; Wang, C.; Ercan, M.; Culver, T.; Beling, P.; Goodall, J. Reinforcement learning-based real-time control of coastal urban stormwater systems to mitigate flooding and improve water quality. *Environmental Science: Water Research & Technology* **2022**, *8*, 2065–2086.
156. Essamlali, I.; Nhaila, H.; El Khaili, M. A new architecture of Low Impact Development (LID)-based stormwater management system through Internet of Things (IoT) and Machine Learning integration. *Case Studies in Chemical and Environmental Engineering* **2024**, *10*, 100942.
157. Li, N.; Ma, J.; Huang, S.; Zhu, H.; Sun, Y.; Hu, M. A comparative study of different deep learning models for land use and land cover mapping of flood detention basin. *IOP Conference Series: Earth and Environmental Science* **2022**, *1087*, 012044.
158. Herath, M.; Jayathilaka, T.; Hoshino, Y.; Rathnayake, U. Deep machine learning-based water level prediction model for Colombo flood detention area. *Applied Sciences* **2023**, *13*, 2194.
159. Li, S.; Kazemi, H.; Rockaway, T. Performance assessment of stormwater GI practices using artificial neural networks. *Science of the total environment* **2019**, *651*, 2811–2819.
160. Al Mehedi, M.; Amur, A.; Metcalf, J.; McGauley, M.; Smith, V.; Wadzuk, B. Predicting the performance of green stormwater infrastructure using multivariate long short-term memory (LSTM) neural network. *Journal of Hydrology* **2023**, *625*, 130076.
161. Yang, Y.; Zhu, D.; Loewen, M.; Ahmed, S.; Zhang, W.; Yan, H.; Duin, B.; Mahmood, K. Evaluation of pollutant removal efficiency of urban stormwater wet ponds and the application of machine learning algorithms. *Science of the Total Environment* **2023**, *905*, 167119.
162. Reshadi, M.; Rezanezhad, F.; Shahvaran, A.; Ghajari, A.; Kaykhosravi, S.; Slowinski, S.; Cappellen, P. Assessment of environmental and socioeconomic drivers of urban stormwater microplastics using machine learning. *Scientific reports* **2025**, *15*, 6299.

163. Hussain, M.; Tayyab, M.; Ullah, K.; Ullah, S.; Rahman, Z.; Zhang, J.; Al-Shaibah, B. Development of a new integrated flood resilience model using machine learning with GIS-based multi-criteria decision analysis. *Urban Climate* **2023**, *50*, 101589.
164. Wang, Y.; Zhang, P.; Xie, Y.; Chen, L.; Cai, Y. Machine learning insights into the evolution of flood Resilience: A synthesized framework study. *Journal of Hydrology* **2024**, *643*, 131991.
165. Satour, N.; Benyacoub, B.; El Moçayd, N.; Ennaimani, Z.; Niazi, S.; Kassou, N.; Kacimi, I. Machine learning enhances flood resilience measurement in a coastal area–Case study of Morocco. *Journal of Environmental Informatics* **2023**, *42*, 53–64.
166. Zhang, W.; Hu, B.; Liu, Y.; Zhang, X.; Li, Z. Urban flood risk assessment through the integration of natural and human resilience based on machine learning models. *Remote Sensing* **2023**, *15*, 3678.
167. Abdel-Mooty, M.; El-Dakhakhni, W.; Coulibaly, P. Data-driven community flood resilience prediction. *Water* **2022**, *14*, 2120.
168. Abdel-Mooty, M.; Yosri, A.; El-Dakhakhni, W.; Coulibaly, P. Community flood resilience categorization framework. *International Journal of Disaster Risk Reduction* **2021**, *61*, 102349.
169. Saravi, S.; Kalawsky, R.; Joannou, D.; Rivas Casado, M.; Fu, G.; Meng, F. Use of artificial intelligence to improve resilience and preparedness against adverse flood events. *Water* **2019**, *11*, 973.
170. Costache, R.; Arabameri, A.; Costache, I.; Crăciun, A.; Pham, B. New machine learning ensemble for flood susceptibility estimation. *Water Resources Management* **2022**, *36*, 4765–4783.
171. Chen, J.; Huang, G.; Chen, W. Towards better flood risk management: Assessing flood risk and investigating the potential mechanism based on machine learning models. *Journal of environmental management* **2021**, *293*, 112810.
172. Shikhteymour, S.; Borji, M.; Bagheri-Gavkosh, M.; Azimi, E.; Collins, T. A novel approach for assessing flood risk with machine learning and multi-criteria decision-making methods. *Applied geography* **2023**, *158*, 103035.
173. Rafiei-Sardooi, E.; Azareh, A.; Choubin, B.; Mosavi, A.; Clague, J.; P., G.; M., M.; M.P., G.; S., R.; K.D..; et al. Evaluating urban flood risk using hybrid method of TOPSIS and machine learning. *International Journal of DDeroliya* **2021**, *851*, 158002.
174. Deroliya, P.; Ghosh, M.; Mohanty, M.; Ghosh, S.; Rao, K.; Karmakar, S. A novel flood risk mapping approach with machine learning considering geomorphic and socio-economic vulnerability dimensions. *Science of the Total Environment* **2022**, *851*, 158002.
175. Al-Kindi, K.; Alabri, Z. Investigating the role of the key conditioning factors in flood susceptibility mapping through machine learning approaches. *Earth Systems and Environment* **2024**, *8*, 63–81.
176. Eini, M.; Kaboli, H.; Rashidian, M.; Hedayat, H. Hazard and vulnerability in urban flood risk mapping: Machine learning techniques and considering the role of urban districts. *International Journal of Disaster Risk Reduction* **2020**, *50*, 101687.
177. Luo, Z.; Tian, J.; Zeng, J.; Pilla, F. Resilient landscape pattern for reducing coastal flood susceptibility. *Science of The Total Environment* **2023**, *856*, 159087.
178. Yuan, F.; Fan, C.; Farahmand, H.; Coleman, N.; Esmalian, A.; Lee, C.; Patrascu, F.; Zhang, C.; Dong, S.; Mostafavi, A. Smart flood resilience: Harnessing community-scale big data for predictive flood risk monitoring, rapid impact assessment, and situational awareness. *Environmental Research: Infrastructure and Sustainability* **2022**, *2*, 025006.
179. Debnath, J.; Sahariah, D.; Mazumdar, M.; Lahon, D.; Meraj, G.; Hashimoto, S.; Kumar, P.; Singh, S.; Kanga, S.; Chand, K.; et al. Evaluating flood susceptibility in the Brahmaputra river basin: An insight into Asia's Eastern Himalayan floodplains using machine learning and multi-criteria decision-making. *Earth Systems and Environment* **2023**, *7*, 733–760.
180. Liu, C.; Mostafavi, A. Floodgenome: Interpretable machine learning for decoding features shaping property flood risk predisposition in cities. *Environmental Research: Infrastructure and Sustainability* **2025**, *5*, 015018.
181. Taromideh, F.; Fazloula, R.; Choubin, B.; Emadi, A.; Berndtsson, R. Urban flood-risk assessment: Integration of decision-making and machine learning. *Sustainability* **2022**, *14*, 4483.
182. Lyu, H.; Yin, Z.; Zhou, A.; Shen, S. MCDM-based flood risk assessment of metro systems in smart city development: A review. *Environmental Impact Assessment Review* **2023**, *101*, 107154.
183. Kumar, R. A Comprehensive Review of MCDM Methods, Applications, and Emerging Trends. *Decision Making Advances* **2025**, *3*, 185–199.
184. Diaconu, D.; Costache, R.; Popa, M. An overview of flood risk analysis methods. *Water* **2021**, *13*, 474.
185. Li, C.; Sun, N.; Lu, Y.; Guo, B.; Wang, Y.; Sun, X.; Yao, Y. Review on urban flood risk assessment. Sustainability. *Sustainability* **2022**, *15*, 765.

 doi:10.20944/preprints202504.1887.v1

186. Fang, L.; Yin, J.; Wang, Y.; Xu, J.; Wang, Y.; Wu, G.; Zeng, Z.; Zhang, X.; Zhang, J.; Meshyk, A. Machine learning and copula-based analysis of past changes in global droughts and socioeconomic exposures. *Journal of Hydrology* **2024**, *628*, 130536.
187. Mekruksavanich, S.; Sooksomsatarn, K.; Jitpattanakul, A. Flooding forecasting system based on water monitoring with IoT technology. In Proceedings of the 2021 IEEE 12th International Conference on Software Engineering and Service Science (ICSESS, 2021, pp. 247–250.
188. Prakash, C.; Barthwal, A.; Acharya, D. FLOODWALL: a real-time flash flood monitoring and forecasting system using IoT. *IEEE Sensors Journal* **2022**, *23*, 787–799.
189. Masoudimoghaddam, M.; Yazdi, J.; Shahsavandi, M. A low-cost ultrasonic sensor for online monitoring of water levels in rivers and channels. *Flow Measurement and Instrumentation* **2025**, *102*, 102777.
190. Jiang, J.; Han, G.; Shu, L.; Guizani, M. Outlier detection approaches based on machine learning in the internet-of-things. *IEEE Wireless Communications* **2020**, *27*, 53–59.
191. Jan, O.; Jo, H.; Jo, R.; Kua, J. Real-time flood monitoring with computer vision through edge computing-based Internet of Things. *Future Internet* **2022**, *14*, 308.