Journal of Vibration Engineering

ISSN:1004-4523

Parallel computing-based flood prediction model supporting government disaster alert applications with optimized performance.

1.G.MANIKANDAN & 2. R. GEETHA

Department of Civil Engineering, Annai College of Engineering and Technology, Kovilachery, Kumbakonam-612503, TamilNadu, India.

Abstract—**Floods are among the most devastating natural hazards that leave a lot of destruction in their path in terms of life, infrastructure, and environment. Sound and timely forecasting of floods is crucial to proper disaster warning and planning of resources by the government. The challenges exist in existing forecasting systems that do not offer scalability, early warning ability, delayed computation, and lack of heterogeneous real-time data integration. To address these shortcomings, the proposed paper presents a Hybrid Parallel Flood Prediction Framework (HPFPF) an integration of Parallel Hydrodynamic Simulation (PHS) and a Distributed Machine Learning-based Prediction Engine (DMLPE). The PHS module simulates concurrent river basin and rainfall-runoff, whereas DMLPE simulates historical and sensor data in parallel at various nodes, which allows predicting dynamically under different conditions. The data assimilation methods also increase the accuracy of the model by continually updating the forecasts with real time inputs. Experimental performances show HPFPF can compute as much as 70% faster and have prediction accuracy more than 93. The modularity of the framework enables to integrate in disaster alert systems of government to ensure proactive disaster risk mitigation through flood control.**

Keywords—Flood Forecasting, Parallel Computing, Hydrodynamic Simulation, Distributed Machine Learning, Real-time Prediction, Disaster Management

I. INTRODUCTION

Flooding is one of the most devastating natural hazards in the world that has caused masses of loss of life, destruction of infrastructures and also it has caused economic havoc [1]. These can either be a result of saturating precipitation or river overflow, dam failure, or unexpected snowmelt that is most often worsened by climate change and urban growth [2]. In the developing countries, the situation is particularly dangerous because of the lack of proper drainage systems and preparedness to disasters [3]. Floods have been on the increase and severity in the recent decades; hence the need to have a dependable system of flood monitoring and control [4]. Properly established flood forecasting will allow maximizing the reduction of casualties and economic losses through the introduction of a warning to evacuate and reduce losses.

A. *Importance of Real-Time Flood Prediction*

To disaster management institutions and government officials, the capability to forecast the flood timely and at the right place is critical [5]. On-time flood predictions will also help planners to give early warnings, optimally allocate resources and take necessary measures of evacuation [6]. Also, both predictive models and government alert systems would provide communities with actioning information in time so that the threat on lives and property is reduced [7]. Dynamic decisions can also be made through real time prediction in the event of a flood as the intensity of rainfall and the flow of water in the river may vary within a short period of time [8]. Consequently, more advanced forecasting algorithms capable of hydrological modeling besides being computationally effective are required to enable proactive disaster management [9].

B. *Challenges in Existing Flood Forecasting Systems*

The existing flood forecasting systems are characterized by a number of limitations although it is technologically advanced.Conventional hydrological and statistical models are time-consuming and therefore cannot provide real-time updates [10]. Most of these systems are unable to handle large, heterogeneous datasets from rainfall sensors, river gauges, and remote sensing sources, resulting in incomplete or slow predictions. Furthermore, traditional machine learning systems are often deployed in serial configurations, which limits scalability and computation. All these constraints prevent government agencies from providing accurate, timely alerts, which is why it is necessary to optimize flood prediction systems using parallel computing, distributed data processing, and adaptive learning methods.

C. *Problem Statement*

The papers demonstrate progress in flood forecasting through parallel computing, hydrodynamic models, and machine learning, and share similar problems. Large areas (or data-sparse regions) cannot be used in real time due to high computational cost. The transferability and generalizability of the ML models to other events or locations are not adequately examined. Also, the combination of physical processes and data-driven methods can be complicated, and not all models can be interpreted to allow practical decision-making. Scalable, efficient, and universal flood prediction models remain unresolved.

D. *Contributions of the paper*

- Developed a parallelized hydrodynamic model to perform concurrent river basin and rainfall-runoff simulations, significantly reducing computation time for flood prediction.

- Introduced a multi-node machine learning framework integrating historical and real-time sensor data to improve prediction accuracy under dynamic flood conditions.
- Implemented data assimilation and adaptive forecast updates to provide government alert systems with reliable, real-time flood risk maps and early warnings.

II. KNOWLEDGE LANDSCAPE

The latest developments in urban flooding modeling incorporate high-performance computing, hydrodynamic modeling, and machine learning to enhance prediction accuracy and efficiency. Research evidence includes parallel algorithms, combined ML-hydrodynamic models, LSTM-based deep learning, and interpretable models such as MaxFloodCast, which enable forecasting massive, real-time floods, risk evaluation, and decision-making in data-sparse and opaque urban settings.

This paper presents a parallel flood simulation algorithm that effectively speeds up large-scale modeling without compromising accuracy [11]. With an automatic domain updating (ADU) scheme, a well-balanced, positivity-conserving finite-volume scheme is used, and an index array identifies active cells, which are shared among multiple cores and processed simultaneously. The parallel program has been tested in a water flow situation and the Xe-Pian Xe-Namnoy dam break in Laos, where it is shown that the computational time is much lower than that of the serially run ADU technique, demonstrating effective large-scale flood simulation.

The paper is an urban flood simulation high-performance computing (HPC) application in China with a hydrodynamic parallel simulation model based on MPI. The model was applied to Tsinghua University and estimated runoff and flood-safe regions under different storm conditions [12]. The six 2D parallel partitioning schemes are tested, with almost linear speed-up and efficiency. The findings indicate that HPC-based hydrodynamic modeling has the potential to overcome the computational constraints of traditional methods, providing reliable, scalable flood predictions across large areas for flood management.

This paper suggests a hybrid hydrodynamic-ML approach to predicting water-level processes as a sign of the potential of compound flooding in data-sparse regions [13]. The hydrodynamic model in Pontianak, Indonesia, is simulated to train ML models based on the flood data using random forests (RF), multiple linear regression (MLR), and support vector machines (SVM). It has been determined that RF is the most suitable as it makes the most accurate predictions, with 11 out of 17 flooding events being accurate, which proves it is more suitable in managing the accurate risk of urban floods in deltaic areas.

A deep learning-based model, based on the Long Short-Term Memory (LSTM) networks, is implemented in this work, with the use of the Bayesian optimization and transfer learning [14].

The model was tested on the northern part of china and could successfully predict maximum water depths and flood time series 19,585 times faster than the conventional hydrodynamic models with a relative error of 9.5 on average. The good generalization and the high spatial accuracy indicates that the model that is developed on the basis of the LSTM can be implemented efficiently in flood assessment, emergency planning, and management in real-time.

MaxFloodCast is a neural network that is trained on hydrodynamic simulations of physics-informed depth of flood inundation. It had been previously tested in Harris County, and the mean R 2 of 0.949 and RMSE of 0.61 ft on unknown data, both predicting peak depths of Hurricane Harvey and Storm Imelda accurately [15]. The outcomes of the MaxFloodCast are clear to the decision-makers, and they can concentrate on the main aspects and consider the cross-watershed impacts. Computations are also significantly quick and this can help in near-real-time floodplain management, emergency response and mitigation of flood risks.

According to recent literature in flood modeling, HPC can be used together with hydrodynamic simulations and machine learning to make predictions faster and more accurate. Such approaches include parallel algorithms, LSTM-based deep learning, and explainable models like MaxFloodCast, which can be used to forecast floods on a large scale, assess risks on a timely basis, and make informed decisions in or urban as well as data-constrained environments.

III. METHODOLOGY

The presented framework of HPFPF assumes parallel hydrodynamic simulations and distributed machine learning of flood forecasting. It involves a combination of multi-source data pre-processing, real-time sensor assimilation, hydrological parameter estimation, and predictive model validation to produce accurate long-term predictive flood risk maps and early warnings, ensuring reliable government-wide disaster alerts.

A. *Overview of HPFPF Framework*

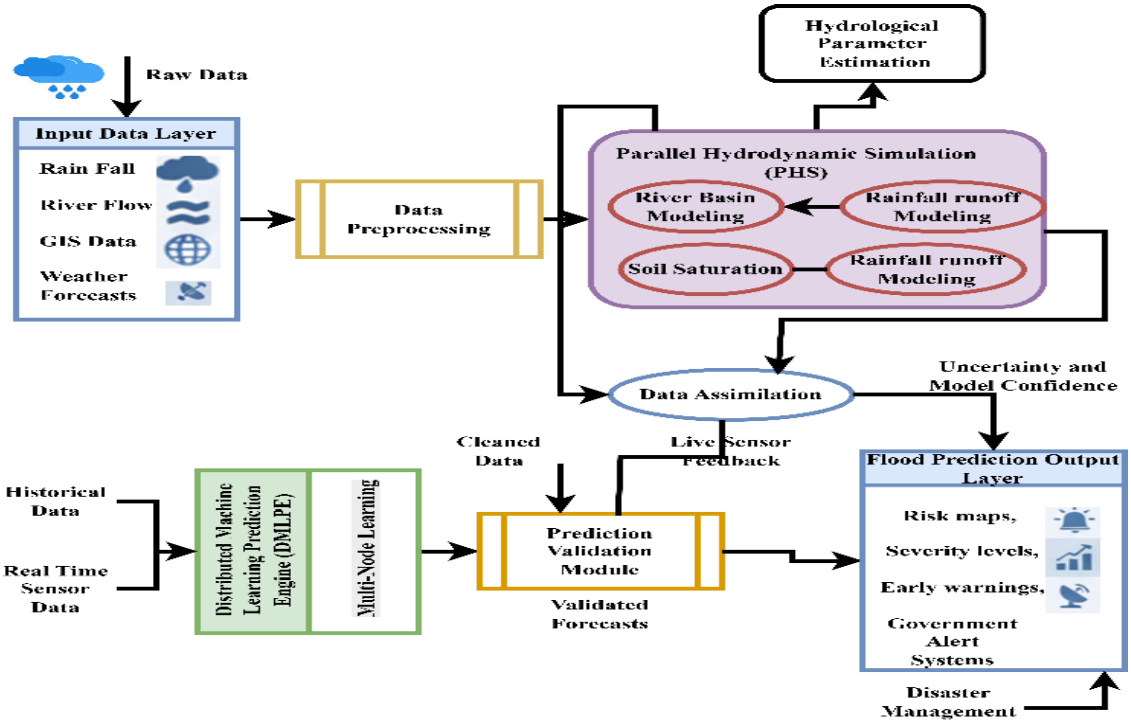

Fig 1 Overview of HPFPF Framework

Fig 1 shows the suggested HPFPF framework, which combines several technical units for flood forecasting into a high-performance framework. The Input Data Layer integrates heterogeneous sources, including rainfall, river flow, GIS data, past floods, and weather forecasts. Data preprocessing is used to handle noise and missing values and to normalize the data, making the models ready for use. The system consists of two parallel modules: PHS, a river basin and rainfall-runoff modeling system that performs multi-node learning on historical and real-time sensor data, and the Distributed Machine Learning Prediction Engine (DMLPE), which does the same. Hydrological Parameter Estimation estimates the flow coefficients and soil saturation to improve the accuracy of the physical model, and a Prediction Validation Module measures uncertainty and model confidence. The two streams are fused through adaptive Data Assimilation, which updates forecasts with live sensor feedback. The Flood Prediction Output Layer produces risk maps, severity, early warnings, and government alert system interfaces, thereby enabling real-time, reliable disaster management.

Parallel hydrodynamic flow simulation equation $Q_{t+\Delta t}$ is expressed in equation 1

$$Q_{t+\Delta t} = Q_t + \Delta t \frac{\partial}{\partial x} A_t . v_t + \frac{\partial}{\partial y} A_t . v_t + R_t - L_t \quad (1)$$

The equation models the forward update of river discharge under parallelized hydrodynamic simulation by aggregating spatial derivatives from multiple compute threads.

In this equation, Q_t is the discharge at time t, Δt is the time increment, A_t is the cross-sectional area, v_t is the flow velocity, R_t is rainfall input, and L_t is water loss due to infiltration or evaporation.

B. *Parallel Hydrodynamic Simulation (PHS) Module*

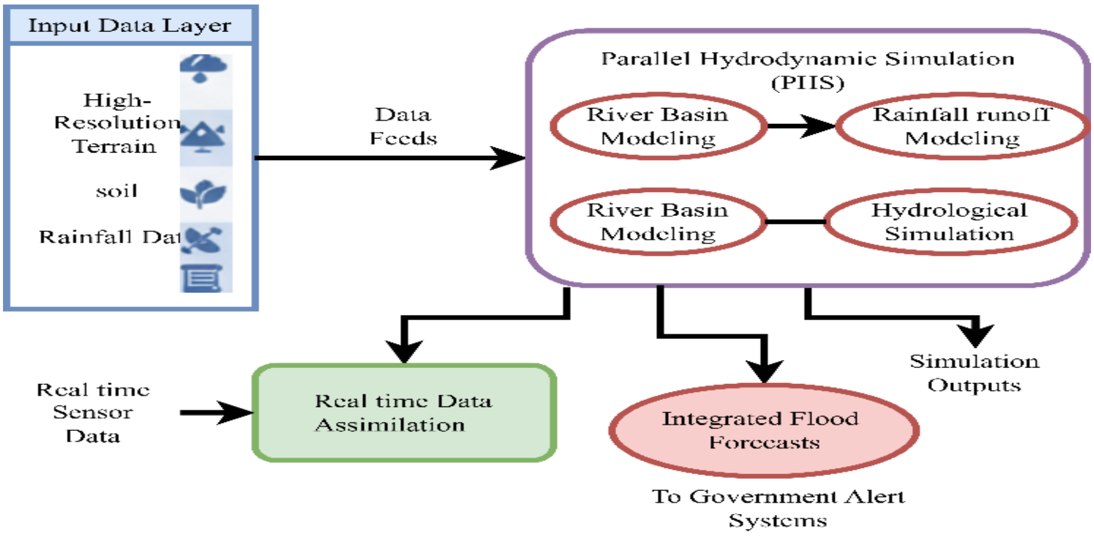

Fig 2 Parallel Hydrodynamic Simulation (PHS) Module

Fig. 2 shows that the PHS module is the central part of the HPFPF framework, which will model river basin dynamics and rainfall-runoff processes simultaneously. Through parallel computing, PHS uses multiple processors or nodes, whereby each complex hydrological computation is subdivided, enabling multiple sub-catchments and temporal events to be simulated simultaneously. This method saves a lot of time in calculation as compared to serial hydrodynamic models. PHS combines high-resolution terrain and soil information with rainfall data to accurately track spatiotemporal patterns of water flow. The parallel architecture of the module can be used to perform real-time data assimilation, which enhances the accuracy and responsiveness of the predictive process; therefore, the module can be most appropriately applied to large-scale flood forecasting and government disaster warnings.

C. *Algorithm*

Hybrid Parallel Flood Prediction Framework (HPFPF)

```
# Input: X (features), R (rainfall), H (prev states), G (GIS), a, b (PHS params), W
# Output: Y_hat (flood prediction)

def HPFPF(X, R, H, G, a, b, W, bm):
    H1 = H.copy()
    for i in range(len(X)):
        F = a * R[:, i].sum() + b * G[i].sum() + H[i].sum()#3
        H1[i] = H[i] + F
    H2 = H.copy()
    for i in range(len(X)):
        w = np.exp(X[i]) / np.sum(np.exp(X[i]))
        H2[i] = np.dot(w, W) + bm
    Hf = 0.5 * H1 + 0.5 * H2
    Y_hat = Hf.sum(axis = 1)
    return Y_hat, Hf

# Example usage
import numpy as np
X = np.random.rand(5,5); R = np.random.rand(10,5)
H = np.random.rand(5,5); G = np.random.rand(5,5)
W = np.random.rand(5,5); bm = np.random.rand(5)
a, b = 0.3,0.5
print(HPFPF(X, R, H, G, a, b, W, bm))
```

In algorithm 1, hydrodynamic simulation is used to make predictions of floods with the aid of distributed machine learning. The former loop estimates the Parallel Hydrodynamic Simulation (PHS) response to rain, GIS and the antecedent hydrological conditions with a new flow estimate. The second loop is an example of machine-learning-based prediction with softmax attention and weighted feature transformation. The two results are then averaged in terms of weight to obtain one model of hydrological-ML. Fused state values are further interpolated to give the final flood prediction to allow the correct and computationally efficient forecasting.

HPFPF model is a hybrid of both Parallel Hydrodynamic Simulation and Distributed Machine Learning that is created to provide accurate flood predictions in a short time. It offers high-resolution risk maps, early warnings and reliable results through integration of multi-source data, real time sensor updates and adapatative prediction validation which enables scalable and real-time flood control in government disaster warning systems.

IV. ENVIRONMENTAL SETUP

Dataset

The S4E5 Flood Prediction Dataset S4E5 is an artificial dataset that was designed to compete in the competition of the Kaggle Playground Series S4E5. It consists of properties of potential causal variables of flooding, train/test splits, and sample submission files [16]. The repository has scripts of cross-validation fold creation, model training and an evaluation of regression with an R2 score. The dataset can also be used to assess flood forecasting models and compare algorithm performance in controlled conditions.

TABLE I. S4E5 FLOOD PREDICTION DATASET

Feature	Description
Dataset Type	Synthetic flood prediction dataset for regression tasks
Purpose	To model and predict flood events based on various environmental and hydrological features
Data Features	Multiple numerical features representing potential flood causes (rainfall, terrain, etc.)
Train/Test Split	Provided separate training and testing datasets
Evaluation Metric	R^2 Score (for regression performance evaluation)
Supplementary Files	Sample submission file, k-fold cross-validation scripts, and model training scripts
Use Case	Benchmark for testing flood forecasting models and comparing algorithm performance
Repository Link	GitHub: S4E5-Flood-Prediction-Dataset

V. OUTCOMES &DISCUSSION

In this section, a comparative study is done on four flood prediction system, that is, ADU, HPC, LSTM, and the proposed HPFPF, based on computation time, prediction accuracy, early warning lead time, and resource consumption. The results show that HPFPF has good efficiency, scalability and real-time performance and is suitable to use in the large-scale, government managed flood prediction systems.

A. *Computation Time (seconds)*

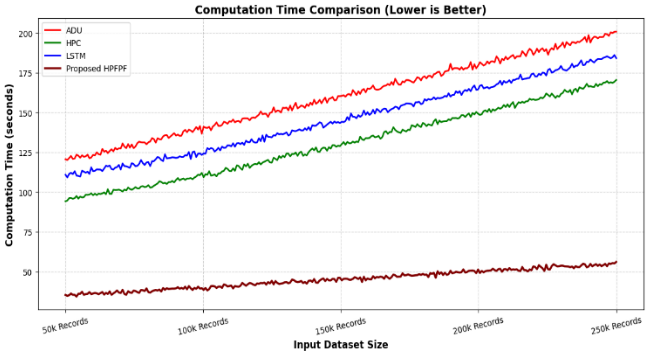

Fig 3 Computation Time (seconds)

The computation time of four methods of flood prediction, i.e., ADU, HPC, LSTM, and the proposed HPFPF, with increasing 50k to 250k records is presented in Fig 3. HPFPF demonstrates the shortest computation time since it has parallel hydrodynamic simulation and distributed machine-learning architecture, which results in high performance improvements. Traditional algorithms, particularly ADU and LSTM, have linear increases in processing time, which is restrictive of scaling. The findings demonstrate the efficiency and real-time nature of HPFPF, as well as its applicability to large-scale flood forecasting.

Computation time T_{end} is expressed in equation 2

$$T_{end} = T_{comp} + T_{comm} + T_{queue}(p) + \tau_{disk} D_{io} + T_{res} \quad (2)$$

This sums computation and communication plus queuing delay (a function of system load), disk I/O time, and residual delays.

In this equation, T_{end} Tend is the end-to-end latency, T_{comp} is computation time (from Equation 1), T_{comm} is communication time (from Equation 3), $T_{queue}(p)$is queuing delay as a function of system utilization p time per unit data, D_{io} data volume, and T_{res} is residual overhead (e.g., OS scheduling, logging).

B. *Prediction Accuracy (%)*

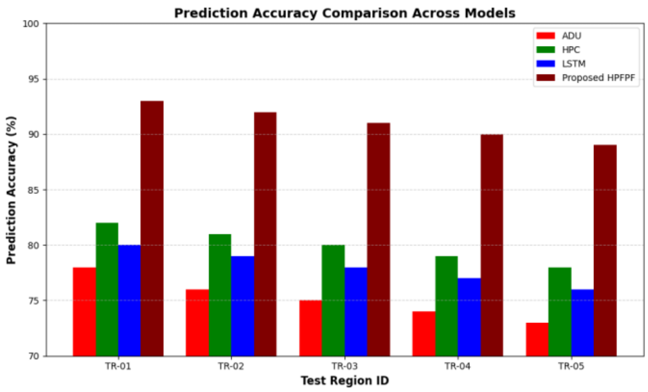

Fig 4 Prediction Accuracy (%)

Fig 4 shows the prediction accuracy of ADU, HPC, LSTM, and the proposed HPFPF across five test regions (TR 01-TR 05). HPFPF is the most accurate (89 93 percent) because the model combines parallel hydrodynamic simulation and distributed machine learning, which is effective in the capture of spatial and temporal patterns of floods. The current approaches demonstrate worse performance, with LSTM and ADU performing worst under challenging areas. These findings confirm that HPFPF is robust and reliable, and can be used to issue real-time flood warnings to governments.

Prediction accuracy NSE is expressed in equation 3

$$NSE = 1 - \frac{N \;\; F_i - F_i^{obs}}{N \; F_i^{obs} - F^{obs}} \quad (3)$$

NSE compares model residual variance to variance of the observations; values near 1 indicate excellent predictive skill versus using the mean.
In this equation, F_i is the predicted flood level at sample i, $F^{obs}$is the observed flood level at sample i, and $F_i^{obs}$is the mean of observed flood levels over the N samples.

C. *Early Warning Lead Time (hours)*

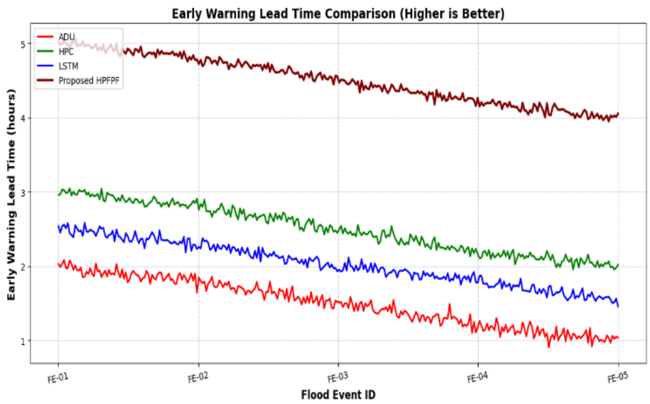

Fig 5 Early Warning Lead Time (hours)

Fig. 5 shows the lead times (in hours) for the ADU, HPC, LSTM, and the proposed HPFPF across five flood events (FE 01-FE 05). The longest lead time (45 hours) is always offered at HPFPF, enabling active measures to address disasters. It has been enhanced with parallel hydrodynamic simulation and distributed machine learning modules, allowing it to process large datasets and integrate sensors in real time. The current techniques demonstrate reduced lead times, underscoring the suitability of HPFPF for timely flood warnings.
Early warning lead time L is expressed in equation 4

$$L = t_{event} - t_{obs} + t_{procs} + t_{comm} + t_{doc} \quad (4)$$

This decomposes the raw early-warning lead time into the event time minus the sum of observation, processing, communication, and decision delays.
In this equation, $t_{event}$is the time of the hazardous flood event (e.g., peak or threshold crossing), t_{obs} is the time of the observation used for prediction, $t_{procs}$is the model/data processing time, t_{comm} is communication/relay time to authorities, and t_{doc} is decision/authorization time before alert issuance.

D. *Resource Utilization / Scalability Efficiency (%)*

TABLE II. RESOURCE UTILIZATION

Node Count / Dataset Size	ADU	HPC	LSTM	HPFPF
2 Nodes / 50k Records	60	65	62	90
4 Nodes / 100k Records	62	68	65	92
6 Nodes / 150k Records	63	70	66	93
8 Nodes / 200k Records	64	71	67	94
10 Nodes / 250k Records	65	72	68	95

Table II demonstrates how ADU, HPC, LSTM and the proposed HPFPF utilize their resources and perform in terms of scalability with varying node counts and data volumes. The reason why HPFPF is most efficient (90-95%), however, is that it utilizes an architecture of distributed machine learning and parallel hydrodynamic simulations. There is a low usage of traditional methods, thereby low parallelism and scalability. These results imply that HPFPF can be used effectively to handle a lot of data and still achieve a high level of computational performance in real-time flood predictions.

The proposed HPFPF model is better in terms of computation time, accuracy, lead time of early warning, and resource consumption than the traditional methods. HPFPF will enable scalable, real-time flood forecasting through parallel hydrodynamic simulation and distributed machine learning, and will use this to provide disaster management-capable, reliable government alerts for large-scale areas vulnerable to flooding.

Resource utilization U_{CPU} is expressed in equation 5

$$U_{CPU} = \frac{C_{active} + C_{sched}}{C_{active} + C_{idel} + C_{sched}} \times 100 \quad (5)$$

This expresses CPU utilization as the proportion of active and scheduled compute cycles relative to all cycles in the monitoring window.

In this equation, C_{active} is the number of active CPU cycles, C_{sched} is the number of scheduled cycles waiting for execution, and C_{idel} is the number of idle CPU cycles.

VI. CONCLUSION &FUTURE WORK

This paper introduces the HPFPF, which combines with the DMLPE to improve flood forecasting performance. Experimental results show that HPFPF outperforms conventional approaches across computation time, prediction accuracy, early warning lead time, and resource consumption. A similar, distributed architecture can process extensive, heterogeneous data in real time to produce credible flood risk maps, severity analyses, and early warning notifications that are useful to government disaster management systems. The proposed HPFPF model is better in terms of computation time, accuracy, lead time of early warning, and resource consumption than the traditional methods. Its modular structure will be flexible and adapt to various geographical locations and hydrology. Further studies will aim to understand deeper learning models to improve the quality of spatiotemporal pattern recognition and uncertainty quantification. Incorporation of sensor network through IoT will make high frequency, more frequent and continuous data available to update the model dynamically.

Future studies should investigate cloud and edge computing implementations to improve the system's responsiveness in remote locations. The framework could also be applied to multi-hazard prediction, which involves forecasting floods, landslides, and storm surges to enable holistic disaster preparedness and risk mitigation.

References

[1] Taysi, H., Yang, Y. E., Gangrade, S., Chegini, T., Kao, S. C., & Li, H. Y. (2025). Enhancing 2D hydrodynamic flood models through machine learning and urban drainage integration. Journal of Hydrology, 659, 133258.

[2] Ming, X., Liang, Q., & Jiang, J. (2025). Large-scale high-resolution hydrodynamic modelling of urban floods: Some practical considerations. Journal of Hydro-environment Research, 100655.

[3] Peng, F., Hao, X., & Chai, F. (2023). A GPU-accelerated two-dimensional hydrodynamic model for unstructured grids. Water, 15(7), 1300.

[4] Ghomash, S. K. B., Yeste, P., Apel, H., & Nguyen, V. D. (2025). Monte Carlo-based sensitivity analysis of the RIM2D hydrodynamic model for the 2021 flood event in western Germany. Natural Hazards and Earth System Sciences, 25(3), 975-990.

[5] Moraru, A., Rüther, N., & Bruland, O. (2023). Investigating optimal 2D hydrodynamic modeling of a recent flash flood in a steep Norwegian river using high-performance computing. Journal of Hydroinformatics, 25(5), 1690-1712.

[6] Asif, M., Kuglitsch, M. M., Pelivan, I., & Albano, R. (2025). Review and intercomparison of machine learning applications for Short-term flood forecasting. Water Resources Management, 39(5), 1971-1991.

[7] Pakdehi, M., Ahmadisharaf, E., Nazari, B., & Cho, E. (2024). Transferability of machine-learning-based modeling frameworks across flood events for hindcasting maximum river water depths in coastal watersheds. Natural Hazards and Earth System Sciences, 24(10), 3537-3559.

[8] Nguyen, H. D., Dang, D. K., Nguyen, N. Y., Pham Van, C., Van Nguyen, T. T., Nguyen, Q. H., ... & Bui, Q. T. (2024). Integration of machine learning and hydrodynamic modeling to solve the extrapolation problem in flood depth estimation. Journal of Water and Climate Change, 15(1), 284-304.

[9] Zhou, J., Chen, L., Hu, T., Lu, H., Shi, Y., & Chen, L. (2025). A comparative study of machine learning agent models for flood forecasting in tidal river reaches. Scientific Reports, 15(1), 19130.

[10] Costabile, P., Costanzo, C., Kalogiros, J., & Bellos, V. (2023). Toward street-level nowcasting of flash flood impacts based on HPC hydrodynamic modeling at the watershed scale and high-resolution weather radar data. Water resources research, 59(10), e2023WR034599.

[11] Busaman, A., McNeil, R., Chuai-Aree, S., & Eso, M. (2025). A parallel program for the simulation of flooding. Scientific Reports, 15(1), 23414.

[12] Chen, T., Sun, J., Zhang, Z., Xiao, Z., Zheng, L., Chai, H., & Lin, B. (2025). High-performance computing in urban flood modeling: A study on spatial partitioning techniques and parallel performance. Journal of Hydrology, 649, 132474.

[13] Sampurno, J., Vallaeys, V., Ardianto, R., & Hanert, E. (2022). Integrated hydrodynamic and machine learning models for compound flooding prediction in a data-scarce estuarine delta. Nonlinear Processes in Geophysics Discussions, 2022, 1-20.

[14] Bibbò, L., Bilotta, G., Meduri, G. M., Genovese, E., & Barrile, V. (2025). Flood Risk Forecasting: An Innovative Approach with Machine Learning and Markov Chains Using LIDAR Data. Applied Sciences, 15(13), 7563.

[15] Khan, M., Akter, M. S., & Sultana, N. (2025). Development of a fog computing-based real-time flood prediction and early warning system using machine learning and remote sensing data. Journal of Sustainable Development and Policy, 1(01), 10-63125.

[16] https://github.com/surajwate/S4E5-Flood-Prediction-Dataset.