Cognitive GeoAI Digital Twin for Disaster Prognostication

Mrs. Ganga Naidu[1], Mukesh M[2], Mukilan K[3], Sai Nithick Roshaan S[4], Vaithyanadhan S G[5]
[1]*Assistant Professor, K. Ramakrishnan College of Technology, Trichy, Tamil Nadu, India*
[2,3,4,5]*Student, Department of Artificial Intelligence and Data Science, K. Ramakrishnan College of Technology, Trichy, Tamil Nadu, India*

***Abstract*—A link between inundation, prediction, road accessibility and an interpretable spatial interface is required for urban flood decision support. This paper aims at proposing a cognitive geospatial artificial intelligence (GeoAI) digital twin for disaster prognostication and its initial realisation in the intended testing area of the Adyar River basin in Chennai. The proposed architecture integrates terrain and road data with a NumPy simulation, Random Forest and long short-term memory forecasting, a FastAPI/WebSocket service to feed a CesiumJS interface, and flood-aware routing. River-connected, elevation-based inundation is implemented in prototype code and dijkstra routing with depth-based road closures are implemented. The demonstrations are not a validated implementation of Adyar, but rather are attached to the discussion of Srirangam. Evaluation will be conducted using a planned approach, which will involve comparing the forecasting models with persistence and linear regression using event-separated data, and evaluate route validity, detour distance, and delivery latency. The trained-model comparison and observed-event validation and latency is pending. The contribution consists of an integrated system design, documentation of a scenario prototype and an evaluation plan for testing. Operational flood prediction is future work.**

***Index Terms*—Adyar basin, digital twin, evacuation routing, flood simulation, GeoAI, long short-term memory, Random Forest.**

I. INTRODUCTION

Information on potential inundation and the potential for movement restriction are needed for flood management in Chennai. Adyar basin is the targeted study area for the FloodSense Digital Twin (FloodSense DT) cognitive geospatial artificial intelligence (GeoAI) system for disaster prognostication. The role of river flow, drainage and surface inundation is highlighted in the regional modelling of the 2005 and 2015 floods [1]. A decision support interface is required to tie these spatial processes to easily understood descriptions of road conditions as well as the limitations of the model.

The project focuses on the feasibility of integrating a light-weight inundation model, machine learning predictions on short-term inundation, flood-aware routing, and 3D visualization using a browser. The main research questions are whether the Random Forest and long short-term memory forecast the flood depth or the risk more accurately than the simple persistence and linear baselines. The questions that help support are related to spatial plausibility, route feasibility under varying flood conditions, and the time required to provide updates to interface.

The system architecture, preliminary implementation and evaluation methodology are presented. The planned Adyar system of research is different from the current scenario prototype. The latter is a demonstration of the components and not an established calibrated flood forecast or an operating evacuation service. The distinction is important because even if the map is well-intentioned and the water levels and access assumptions are not valid, the map can look convincing.

II. RELATED WORK AND RESEARCH POSITION

Eleven studies were reviewed (and some method papers for the foundations) in a pilot literature review from 2023-2026. The search themes include: Flood forecasting; Digital twins; Regional hydrodynamic modelling; Evacuation routing. The review inspires integration between these functions.

It does not constitute an absolute refusal that there may be similar systems and the proposed contribution ought to be evaluated in the context of the systems already presented in the reviewed literature and not as a novelty claim universal.

Kumar and his colleagues coupled models of river, sewer, and surface-flow with rainfall-runoff modelling and high-resolution topography to model the Adyar basin. Their application of flood depths that were measured is an appropriate example of the validation of regional inundation modelling. The simplified prototype that is

being studied here cannot take on that study's accuracy. It may be used to check the plausibility of published spatial patterns, or a quantitative comparison must be based on compatible terrain, event input and observations.

In order to predict the flood in time and space, Zhou and co-workers proposed a deep-learning method based on long short-term memory, Bayesian optimization and transfer learning [2]. Their work is used to investigate fast data-driven surrogates, and is done with an emphasis on the origin of the training targets: agreement with simulated targets translates to reproducing the simulator, but predictive skill with real events requires independent observations. The present study therefore makes a distinction between simulation evaluation, and observational validation.

Li et al. integrated a two-dimensional hydrodynamic model, road-section weighting and Dijkstra routing for flood-avoidance [3]. This defines a direct relationship between inundation and transport decisions. FloodSense DT suggests to integrate a decision layer with evolving predictions and a live spatial interface. It will not be based on the software libraries, but on the reproducible comparison to the baseline and evaluation of integrated behaviour.

III. PROPOSED SYSTEM AND DATA

A. Study area description

1. Study area description or mapping

Proposed evaluation area is the Adyar river basin at Chennai. The Shuttle Radar Topography Mission (SRTM) elevation data [4] is used to prepare terrain and OpenStreetMap data is used for roads and buildings. Road network acquisition and preparation will be done with the planned tool OSMnx [5]. The chosen length of the approximately 30m terrain for the system is suitable for a very simple basin representation, but may not be able to represent small drains, embankments, and local depressions that have a significant impact on urban flooding.

Data from terrain, road network have been gathered, the rainfall source and access arrangements are yet to be settled. Where available, historical rainfall from the India Meteorological Department will be proposed otherwise the explicit labelled design storm will be used. Therefore, at this time, a complete dataset of rainfall and flood depths is not observed. All eventual datasets should contain the date of the download, spatial extent, coordinate reference system, vertical datum, licence, and the history of how the data has been preprocessed.

B. Six-layer architecture

Terrain, roads, buildings and rainfall/scenario inputs are stored in the data layer. Spatial depth states are created by a simulation layer. The purpose of a prediction layer is to forecast the future risk or depth, based on the Random Forest and long short-term memory models. A decision layer is used to transform depth states to road restrictions and to calculate the routes. The service layer consists of fastAPI endpoints and a Websocket channel and the presentation layer is of the terrain, inundation, building exposure, and routes, which are provided by CesiumJS.

The suggested workflow takes inputs, creates a depth state, builds forecast features, forecasts a future state, updates road closures, calculates a route, and sends the output. The current design is based on a simulated or replayed feed, not connection to a live government sensor network is outside of the stated scope. Some other exclusions are a full hydrodynamic solver, structural or dam-break analysis, as well as mobile or short message service (SMS) alert delivery. The intended integration is referred to as a digital twin, but there is not yet a continuous synchronization with a measured physical basin.

C. Predicted outcomes of a plan

Random Forest is one of the tabular learning method that uses ensembles of decision trees [6]. Terrain elevation, slope, flow accumulation, river distance, windows of cumulative rainfall and previous depth are features proposed to be included per cell. The

planned inputs are rainfall and depth sequences, and this is done with a recurrent approach that uses a long short-term memory network [7]. The comparison should be done between two models with same target definition, forecast horizon and held-out events to make a useful comparison.

The project has yet to finalize a comprehensive training set, sequence length, forecast lead time, and model configuration. Prior to training and evaluation, these settings will be specified. If persistence is used, the most recent depth will be a good predictor, and linear regression will give a simple learned comparator. The continuous depth forecasting and risk classification should be assessed individually as an average depth error of low value does not automatically ensure good detection of hazardous road conditions.

IV. PRELIMINARY IMPLEMENTATION

A. Flood protection for a single point, prescribed stage inundation

This is a preliminary solution which relies on a grid of terrain (NumPy) and a four-neighbour breadth-first search starting from mapped river cells. Cells can be inundated when their terrain elevation is below the water-surface elevation. The cells that can be reached from a river seed are searched and the depth of the cells is equal to the difference between water surface and terrain elevation. If a cell is not connected to any cell with a value, it is assigned a value of zero. Four-neighbour traversal does not allow for immediate diagonal connections between neighbors (grid cells) that are blocked by other space-filling elements.

The code employs an illustrative base water-surface elevation of 69.5 m, and may be applied to a user-supplied peak rise with a sinusoidal scenario trajectory. Without an explicit rise, legacy presets relate the rainfall to stage by using an uncalibrated relationship. The baseline is NOT a river gauge measurement. The scenario frames are at 0, 3, 6, 12, 18 and 24 hours, and are scenario positions rather than validated arrival times. The implementation does not address momentum or water-volume conservation, and should be called connected stage-based inundation, not propagation of rainfall-runoff as proposed.

The prototype transforms connected depths into severity zones mapped to them and calculates building exposure based on maximum sampled depths in each building footprint. Its depth bands use boundaries of 0.05, 0.5, and
1.5 m. These are not validated damage functions but are categories for implementation of the damage. The reported risk index is not a probability of flooding or a calibrated prediction of human or financial loss, it is a percentage of the number of mapped buildings that are classified as affected.

B. A* – implemented routing and planned A* extension.

The routing appendix creates a walking graph based on road features mapped by OpenStreetMap, and filters out certain inaccessible roads, e.g., because of terrain, and links together road segments with common OpenStreetMap vertices. Samples terrain cells along segments at about 8 m intervals. If the sampled depth is greater than 0.15 m or if there is a terrain error, a segment will be removed. When the underlying grid is flooded, then the deck elevations of bridges with unknown elevations are conservatively closed. It is a prototype rule and will need to be assessed in the context before being used in practice.

The priority queue implemented is Dijkstra routing because it does not use a heuristic to accumulate path length. A* routing is suggested for the latter system with a distance heuristic [8] which is not illustrated in the appendix. There may be endpoints rejected by the current code that are >250 m away from mapped accessible roads and there may be a no-available-route response if closures or network disconnections prevent a route. The return trip is only possible under the assumptions that are encoded: The endpoint access is not yet verified, the shelter is suitable and the destination is safe.

C. Evidence of demonstration and completion

The prototype demonstrations are available for buildings locations in Srirangam along rivers Kaveri and Kollidam, roads and inundation overlay, and building-depth inspection. They also demonstrate a prescribed water-level-rise scenario and also a no-available-route display. These images are a partial example of the interface and scenario demonstration to be performed at a different site to the proposed Adyar study. They are not validation for an observed Adyar flood, for the trained forecasting models, or

for the reliability of directions.

Completed work encompasses the research definition, pilot literature review, architecture, selection of technology, preparation of repository, and evaluation plan. Before any integration can take place, preliminary simulation and routing components are provided. Further necessary processes, such as complete adyar preprocessing, trained-model comparison, flood-aware a* integration, observed-event validation, and latency testing, need to be established. The available evidence thus justifies the description of implementation and proposed method, but not the experimental forecasting performance.

V. PLANNED EVALUATION AND LIMITATIONS

A. Data validation and data preparation processes

The planned evaluation will be harmonized with terrain, roads and rainfall and will be based on a clearly documented spatial and temporal reference, where a definite terrain and road will be defined, and rainfall will be simulated. For training the earlier storms or periods of time will be used and later periods of time will be used for testing; if data allows, a whole storm will be withheld. The similarity of adjacent time steps in the same scenario should not be spread randomly over training and testing as this could be seen as over generalization. The parameters for pre-processing should be optimized with the training data only.

The partitions used for the random forest tuning will be the same as the temporal or event grouping. The performance of long short term memory training will be repeated for a number of random seeds, and the mean performance as well as the spread will be reported. The final test set needs to be examined after model selection. If only scenario targets are available, the study will report surrogate-model skill against scenario targets and leave claims regarding actual flood forecasting for when independent event observations come in.

B. Component and integrated measures

The forecast error of depth is evaluated by root mean square error (RMSE), mean absolute error (MAE) and coefficient of determination (R^2). A hydrological skill measure will be given by the Nash–Sutcliffe efficiency (NSE) measure against a mean-observation predictor. Precision, recall, F1 and accuracy will be employed to classify risk, with special emphasis on the number of missed hazardous cases. No model is assumed to be better than any other model before the evaluation and results will be compared on the same test partitions for both persistence and linear regression.

The extent and depth will be validated with documented flood information, including the 2015 flood event when available, through simulation. The similarity between images will be used as a plausibility check. For returned paths, routing will be evaluated for avoiding all closed segments, path length compared to the unconstrained shortest path, and computation time. Disconnected-network and no-route cases should not be deleted, but left in the assessment. A graph-valid path is NOT an evacuation safety.

Integrated testing will assess the end-to-end update latency and browser frame rates for a specific grid size, road-graph size, hardware configuration and frequency of updates. Updates within 1 second are a "best guess" rather than a measure of the actual performance. Forecast skill, route computation time and interface responsiveness needs to be reported separately, otherwise the updates to the interface can be wrongly interpreted as an accurate forecast.

C. Conclusions and reproducibility

Conclusions are limited by coarse terrain, road attributes that are incomplete, uncertain bridge elevations, and a lack of rainfall and observed-depth records. A stage only connectivity model does not include drainage hydraulics, flow velocity, surveyed barriers, or bathymetry. The path of its water levels is simple and does not set a time or level of flooding. Using this output for model training can perpetuate these simplifications. New spatial inputs and location-specific validation is required for transfer from Srirangam to Adyar.

Reproducibility will hinge on versioned codes, fixed random seeds, a dependency specification, data downloader or preparation scripts, and a record of all inputs and settings for each experiment. Redistributed data should be accompanied by OpenStreetMap attribution and applicable source licences. The interface should be clear and convey the outputs of a scenario and the status of the decisions made from it

as a decision support tool. With the data available, zone-level error analysis is suggested, as the level of completeness in the map and the quality of terrain can differ from area to area. Beyond the current prototype, verified access conditions and destinations in the field are required for use.

VI. CONCLUSION

This work designs an integrated flood decision-support system and records the preliminary in-connect and in-road parts of it. The planned research is an Adyar-basin comparison of Random Forest and LSTM forecasting methods with persistence and linear baselines. There are indications that support scenario exploration and system development, but still needs to be validated the forecasting accuracy, hydrological validity, A* integration and real-time performance. There are still challenges to be addressed, such as data provenance, regional implementation, and component and end-to-end evaluation, before the operational claims can be made.

REFERENCES

[1] Ajith Kumar K et al., "Monitoring and simulation of Chennai floods through a high-resolution topographic dataset," Discover Cities, vol. 3, art. 62, 2026, doi: 10.1007/s44327-026-00238-8.

[2] Q. Zhou et al., "A deep-learning-technique-based data-driven model for accurate and rapid flood predictions in temporal and spatial dimensions," Hydrology and Earth System Sciences, vol. 27, pp. 1791–1808, 2023, doi: 10.5194/hess-27-1791-2023.

[3] B. Li et al., "High-resolution flood numerical model and Dijkstra algorithm based risk avoidance routes planning," Water Resources Management, vol. 37, pp. 3243–3258, 2023, doi: 10.1007/s11269-023-03500-5.

[4] T. G. Farr et al., "The Shuttle Radar Topography Mission," Reviews of Geophysics, vol. 45, no. 2, RG2004, 2007, doi: 10.1029/2005RG000183.

[5] G. Boeing, "OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks," Computers, Environment and Urban Systems, vol. 65, pp. 126–139, 2017, doi: 10.1016/j.compenvurbsys.2017.05.004.

[6] L. Breiman, "Random forests," Machine Learning, vol. 45, no.1, pp.5–32,2001, doi: 10.1023/A:1010933404324.

[7] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.

[8] P. E. Hart, N. J. Nilsson, and B. Raphael, "A formal basis for the heuristic determination of minimum cost paths," IEEE Transactions on Systems Science and Cybernetics, vol. 4, no. 2, pp. 100–107, 1968, doi: 10.1109/TSSC.1968.300136.