Deep Learning Approach to Predict Peak Floods and Evaluation of Socioeconomic Vulnerability to Flood Events: A Case Study in Baltimore, MD

Ruoyu Zhang
School of Data Science
Department of Environmental Sciences
University of Virginia
Charlottesville, VA
rz3jr@virginia.edu

Hyunglok Kim
School of Data Science
Department of Engineering Systems and Environment
University of Virginia
Charlottesville, VA
hk5kp@virginia.edu

Emily Lien
School of Data Science
University of Virginia
Charlottesville, VA
egl6a@virginia.edu

Diyu Zheng
School of Data Science
University of Virginia
Charlottesville, VA
dz2fc@virginia.edu

Lawrence Band
Department of Environmental Sciences
University of Virginia
Charlottesville, VA
leb3t@virginia.edu

Venkataraman Lakshmi
Department of Engineering Systems and Environment
University of Virginia
Charlottesville, VA
vl9tn@virginia.edu

***Abstract*—As the intensity and frequency of storm events are projected to increase due to climate change, local agencies urgently need a timely and reliable framework for flood forecasting, which can downscale from watershed to street level in urban areas. Integrated flood forecasting with property data, the flood prediction provides further insight into long-term flood risk at household level, which helps the future research of environmental justice. This study uses deep learning (DL) methods and integrated radar-based rainfall data to quickly predict the peak flood quantity and the time of peak flood for each storm event, and we implement an analysis of the property and demographic data with respect to stream proximity will provide a way to quantify economics. Our model is capable of predicting peak stages reasonably well ($R^2 > 0.8$) at two of our study watersheds, but the prediction timing of peak flood is not responding to our predictors. For the flood risk analysis, there is evidence that there is socioeconomic discrimination in those who are at a higher risk of being flooded. This study is a beginning of future research on flood prediction and risk assessment, our current predictors and model structure need further improvement to accurately predict short-term flood quantity and timing in the future research.**

***Index Terms*—ANN, flood forecasting, flood risk, socioeconomic effects**

I. Introduction

As both the intensity and frequency of storm events are projected to increase due to climate change [1], local agencies are in urgent need of accurate flood forecasting so that they can better protect people's lives and infrastructures in urban areas. With access to better quality hydrologic, topographic, and meteorological data which have much finer spatial and temporal resolutions than ever, we are able to improve flood forecasting with more accurate estimations of the peak flood quantity and the time to peak flood from the initiation of a storm event. The increasing coverage of 1-meter Lidar digital elevation model (DEM) in the United States enables detailed delineation of stream geomorphology; Hourly hydrologic observations (e.g., stream discharge and stage) at United States Geological Survey (USGS) gages provide sufficient data to analyze the patterns of flood events at multiple locations that can potentially promote flood forecasting in the future. Many studies [2]–[5] have used machine learning methods to predict daily discharge worldwide, proving their high-accuracy on predicting hydrological processes in the future. However, we do not find studies predict hydrological behaviors at sub-hourly scale though we have the sub-hourly data for several years.

Recently, a new framework, GeoFlood [6], was proposed to map inundated areas by constructing synthetic rating curves from high-resolution DEM. Coupling the GeoFlood with the National Water Model (NWM) which forecasts discharge quantity, we can estimate the water level of a stream from its rating curve and map inundation areas based on the DEM data. However, two major limitations in the GeoFlood may undermine the accuracy of flood forecasting and inundation mapping. The first issue arises from the process of constructing synthetic rating curves. An empirical equation called Manning's equation is widely used to estimate discharge, and one of the most important constants in this equation is Manning's roughness coefficient, n. Currently, the n value is assumed to be 0.5 for all streams in the GeoFlood, but the n value varies in magnitude among streams based on channel morphology, bedforms, vegetation, and other factors. At streams without observations, we do not know the correct value of n and cannot construct an accurate rating curve (i.e., the relationship between the discharge and water stages) to forecast floods and inundation areas. The second issue is the quality of discharge forecasting from the NWM, because the simulated discharge may not consider urban storm sewers and other stormwater management facilities and the accuracy of discharge is unguaranteed. Eventually, any overestimation or underestimation of discharge can disrupt the accuracy of inundation mapping.

Here, we propose a machine learning framework to predict short-term flood quantity and timing, and we also analyze the long-term flood risk at household level in this study. (integrate better) Machine learning is one of the most revolutionary developments in data science, and it provides a new framework for solving short-term flood forecasting. With sub-hourly observation of rainfall and discharge in urban watersheds, we can understand how the flood quantity and timing is related to rainfall amount and intensity and forecasting future flood severity with weather forecast at sub-hourly scale. In this study, we start from data reached from the densely populated Baltimore region, where we have high-quality rainfall and discharge data for about a decade. When data becomes available on greater spatial coverage, we can potentially extend our method to the continental United States and other places in the world.

Flood hazard prediction [7], [8], machine learning based damage evaluation [9], and environmental justice with respect to pollution [10], [11] have all been studied, however it is difficult to find literature outlining whether or not there is discrimination in American cities with respect to flood risk in long-term. Therefore, building on the framework of the environmental inequality study by Clark et al., 2014 [11], we propose applying the same approach to measuring flood risk with respect to income level and race in the Baltimore area. This is done with the intention of providing a new depth of information to the flood prediction framework outlined above, so that continued patterns of inequality may be better highlighted over time in future work.

In summary, this project has two objectives:

- Develop a machine learning model for short-term flood forecasting in urban watersheds.
- Analyze the relationship between flood areas and local demographics, to highlight significant connections between income, race, and flood risk (if any exist), and to construct a visualization that could work in conjunction with the flood forecasting to provide ongoing analysis of who is most at risk of being flooded and the property value at risk from flooding.

II. Methods

A. Study Area

We select the Dead Run and Baisman Run that have decades of sub-hourly discharge and rainfall data, yet the Dead Run is a heavily urbanized watershed and the Baisman Run is mostly forested (Fig. 1). The impervious areas cover 37% of the watershed in Dead Run, and forest areas cover 80% in Baisman Run.

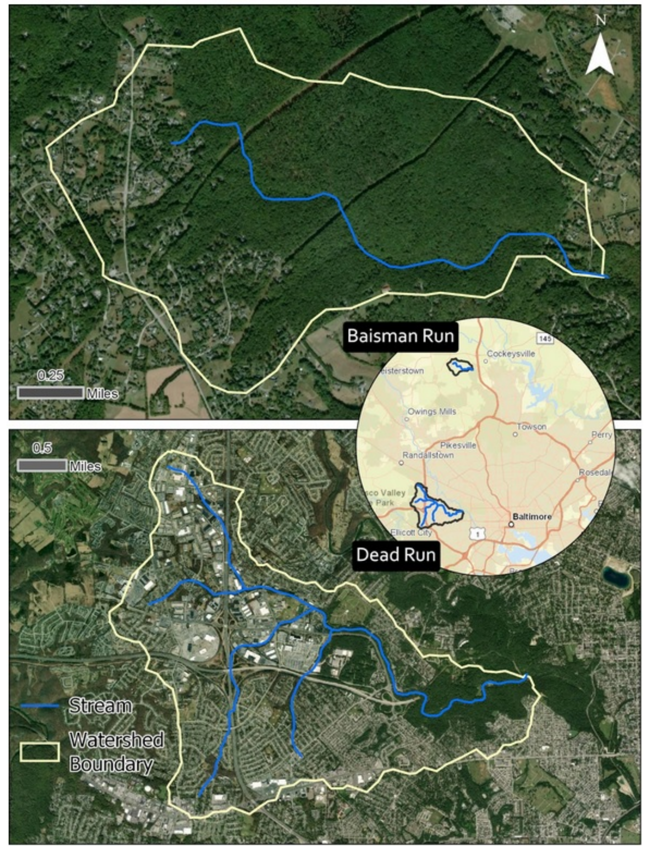

Fig. 1. Two research watersheds, Baisman Run (upper) and Dead Run (lower), located in Baltimore, MD. Two watersheds have significantly different land cover types, where Baisman Run is dominated by forest and Dead Run by urban impervious areas. The watershed area of Baisman Run is 4 km2, much smaller than Dead Run's 20km2. Note we use different scales to map two watersheds

B. Data

1) Radar Rainfall: For various types of hydrological applications, the temporal and spatial resolution of rainfall data is needed. The recent radar equipment enables us to achieve this requirement by providing much finer spatial and temporal scales of rainfall data. Specifically, both precipitation and wind can be measured by NEXRAD (Next Generation Radar). The radar sends out a small pulse of energy, which is dispersed in all directions whenever it hits an object (raindrop, snowflake, moth, bird, etc.). A small portion of that scattered energy is directed back toward the radar, and this information can be used to estimate rainfall intensity [12]. Over Baltimore County and City, Hydro-NEXRAD-2 (HNX2) provides a 1-km resolution rainfall dataset at 15-minute intervals from April to October from 2012 to the present. In this study, we used HNX2 from April 2012 to September 2020 [13]. HNX2 was developed to provide the hydrologic research community with quick access to personalized radar-rainfall items in near real-time. Within a watershed, the rain gage records precipitation only at a point location, whereas radar data has greater spatial coverage for the whole watershed and counts the rainfall cannot be captured by the rain gage.

2) USGS Streamflow: The U.S. Geological Survey (USGS) has established a dense network for monitoring streamflow conditions (e.g., stage, discharge, water temperature, etc.) for the United States. In Baltimore, the USGS keeps observing streamflow conditions at several watersheds where the gradients of watershed size, land cover, and demographic characteristics are quite large. Most USGS gages record water stage and discharge every 15 minutes, except for five gages in Dead Run with 5-minute intervals. The fine-temporal streamflow observations become available starting from various dates at different gages, and in each gage, there are certain periods that data is missing or under revision. For training data preparation, we only consider the periods when both radar rainfall and USGS observations are available and exclude periods that data is missing or under revision.

3) Demographic and Socioeconomic Data: To add descriptive dimensions to the areas of flood study, we collected data on Baltimore City and County from EnviroAtlas. The data from EnviroAtlas has community-based collection practices and includes income data, floodplain land data, population living within a floodplain, and impervious surface land cover data measured at the census block group data level. This data is used to create a more in-depth understanding of who is being flooded, and to analyze if there are any areas/demographics that are statistically more at-risk of being in flood hazard areas. From Census Reporter, we gathered racial demographic data for the City and County at the census block group level to continue the study of those who are being flooded, and whether or not any racial demographics are statistically more likely to live in areas prone to flooding. From the Maryland Department of Planning, we gathered data on current parcel value for a number of buildings in Baltimore City and County based on recent sales. For study purposes, we have aggregated this data to the census block group level so as to better analyze its relationship to other dimensions such as race, income, and floodplain location. The parcel point data was also used to calculate additional information on stream proximity.

C. Data Processing

1) Flood Forecasting Data: In order to produce rainfall data that represent the entire watersheds, we aggregat the 15-minute interval radar rainfall within our studied watersheds for our model training. For Dead Run (USGS 01589330) watershed and Baisman Run (USGS 01583580) watershed, 22 and 5 grids (1-km by 1-km) of radar rainfall data were used to calculate the aggregated intensity of rainfall, respectively. We identify storm events within the 12-hour window since April 2012, when the HydroNexrad becomes available. Next, we find the greatest stage level within the 12-hour window of each storm event. Then, we search the storm event before the peak stage time and identify its start and end time. From the extracted rainfall and stage time-series data. We find no case that the stage peaks when rainfall is still intensifying, which means in our studied gages, peak stage always occurs after the peak rainfall. Finally, we extract the following six predictors: 1) prior peak rainfall duration, 2) prior peak rainfall amount, 3) peak rainfall amount, 4) post peak rainfall duration, 5) post-peak rainfall amount, 6) initial stage level at the start of a storm event. Predictors 1-5 describe the rainfall distribution of a storm event before and after the peak rainfall, and different rainfall distribution patterns may result in various responses of the lag between peak rainfall and stage.

2) Demographic and Socioeconomic Data: From the parcel point data, we were able to calculate a 15m buffer around each parcel point to better include some of the surrounding land around the houses. Using these buffers, we calculated the average height above nearest drainage (HAND [14]) and average horizontal distance to the nearest stream within the 15m buffer. These average HAND and horizontal distance values were then used in our socioeconomic analysis. To better analyze the relationships between socioeconomic, racial demographics, and physical factors, all the data needed to be aggregated at the same level of detail. Census block group was the most common level of measurement, with all the data from EnviroAtlas and Census Reporter gathered at the block group level. The parcel sale price data contained the census block group for each datapoint, so after calculating the HAND and horizontal distances for each parcel sale point these, as well as the current total value, current improvement value, current land value, and square footage, were all aggregated as mean values at the block group level. It is important to note that after aggregation, not all information was available for every block group, which was taken into consideration when assessing the results of our analysis.

D. Flood Forecasting Model Building

We trained forecasting models using six predictors and two response variables described in Section 4.2. Then, the peak stage level and the time-lag to peak stage after peak rainfall will be predicted using 15-min radar rainfall data. We use Multilayer Perceptron (MLP, [15]) to predict the two response variables. We examined two model architectures 1) single-output regression of one response variable with the same 5-layer model structure and 64, 32, 16 hidden layer nodes, respectively and 2) multi-output regression of two response variables with the identical model design, except the output layer contains two nodes. Specifically, the first scenario is training models considering the both response variables (i.e., peak stage level and time-lag to peak stage) simultaneously assuming that two response variables are correlated each other. Second, we trained two models individually for each response variable. For the model training process, we use 80% of our data for training, 10% for validation. The remaining 10% data is used for testing model performance.

E. Flood Risk Analysis

1) Analyzing Relationship between Price and Stream Proximity: To start our analysis, we want to determine if there was a measurable relationship between the height above the nearest drainage, horizontal distance to the nearest drainage, and property value. After an initial assessment of potential relationships through scatterplots, we decided to use linear

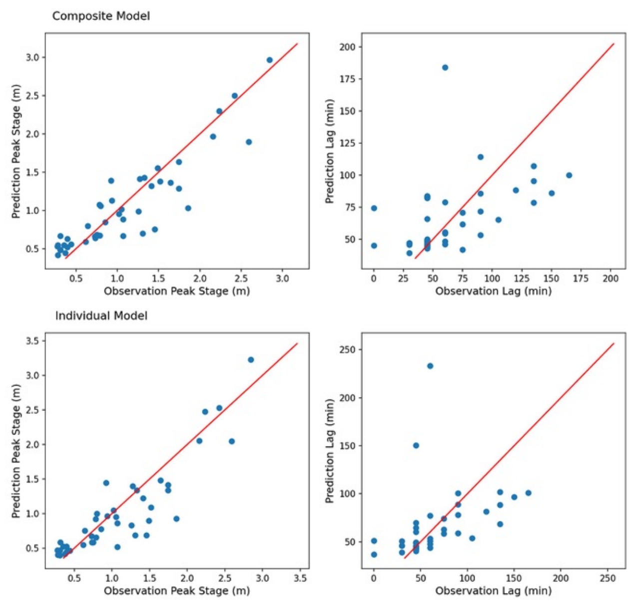

Fig. 2. Comparison of model prediction and observation on peak stage (m, left) and lag (min, right) at Dead Run for composite (upper) and individual (lower) models. Compared to the testing observation data, for the composite model, the prediction of stage has R2 (RMSE) value 0.816 (0.291 m) and the prediction of lag time has R2 (RMSE) 0.231 (33.941 min); For the individual models, the prediction of stage has R2 (RMSE) value 0.790 (0.322 m) and the prediction of lag time has R2 (RMSE) 0.147 (39.970 min)

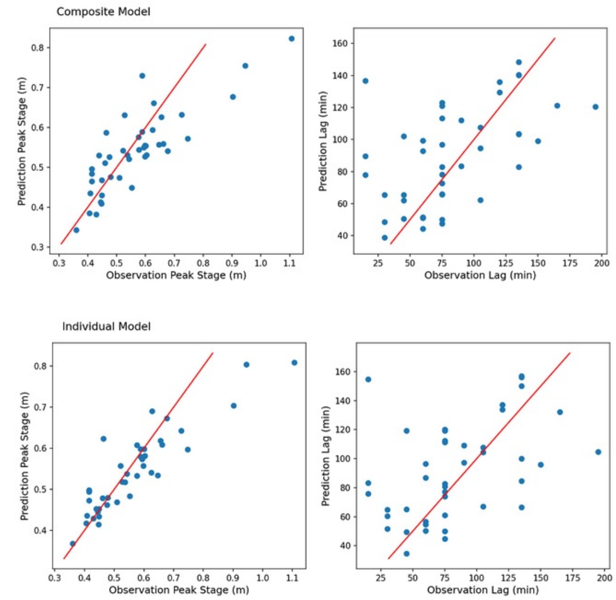

Fig. 3. Comparison of model prediction and observation on peak stage (m, left) and lag (min, right) at Baisman Run for composite (upper) and individual (lower) models. Compared to the testing observation data, for the composite model, the prediction of stage has R2 (RMSE) value 0.691 (0.092 m) and the prediction of lag time has R2 (RMSE) 0.283 (37.45 min); For the individual models, the prediction of stage has R2 (RMSE) value 0.804 (0.079 m) and the prediction of lag time has R2 (RMSE) 0.203 (41.232 min)

regression model to study the relationship between HAND and price per square foot, and segmented regression for the relationship between horizontal distance and price per square foot. Additionally, a multiple linear regression model was created to assess whether HAND and horizontal distance were useful as joint predictors for the average price per square foot.

2) Socioeconomic Disparity in Flood Hazard Zones: An important aspect of our study is identifying whether or not there is discrimination in who is more at risk for flooding based on wealth and race. To answer this question, a series of ANOVA F-tests were conducted on the means of the number of people living in a 100-year flood event zone by income bracket and by race. The null hypothesis to be tested was that the average number of people living in 100-year flood event zones was the same among all income brackets and among all races. To test this, we calculated what we estimated to be the average number of people in each block group who live in a flood hazard zone broken down by income bracket as well as race. For this analysis, an assumption was made that the distribution of the population in a block group that lives in a flood plain is equal to the distribution of income levels and race, and that calculating the populations in this manner would provide a reasonable estimate of the overlap of these populations.

III. Results and Discussion

A. Prediction Accuracy for Peak Stage and Lag

We evaluated the two different model designs at two watersheds in Baltimore, and the predictions of peak stage and lag are plotted. For the models predicting one variable at a time, the predictions for the peak stage showed R2 (RMSE) values at Dead Run (Fig. 2) and Basiman Run (Fig. 3) are 0.790 (0.322 m) and 0.804 (0.079 m), respectively. As the watershed size and discharge of Baisman Run is smaller than the Dead Run watershed, the Basiman Run's peak stage RMSE is much smaller than the Dead Run. For the lag time to peak predictions, both watersheds showed poor predictions than the peak stage predictions – R2 (RMSE) values for Dead Run and Baisman Run are 0.231 (33.94 min), and 0.203 (41.23 min), respectively.

Compared to individual models, the composite model at Dead Run has better performance on both peak stage and lag while the composite model at Baisman Run has worse performance on peak stage but better performance on lag time. Overall, given the rainfall pattern of a storm event, our model generally can predict the peak stage relatively well, which indicates that our method to describe the rainfall amount and intensity from time-series is useful for emergency response teams to estimate the greatest flood stage or quantity from the weather forecast. However, we find the lag time is hard to predict using only rainfall patterns over the whole watershed. The location of center and moving direction of storms can affect the travel time of runoff to the gage and is the factor we do not contain in our predictors. For example, if storms occur near the gage, the stage level at the gage would rise quickly and the lag between peak rainfall and stage should be small; if identical storms occur at up-stream regions, it will take a long time for the water to travel to the gage, therefore,

TABLE I
COMPARISON BETWEEN PERFORMANCE OF INDIVIDUAL AND COMPOSITE MODEL AT DEAD RUN AND BAISMAN RUN

	Dead Run			
	Individual Model		Composite Model	
	R^2	$RMSE$	R^2	$RMSE$
Stage (m)	0.790	0.322	0.816	0.291
Lag Time (min)	0.147	39.970	0.231	33.941
	Baisman Run			
Stage (m)	0.804	0.079	0.691	0.092
Lag Time (min)	0.203	41.232	0.283	37.450

TABLE II
POPULATION WEIGHTED AVERAGE INDICATOR OF DIFFERENT INCOME GROUPS

Income Group	Population in Floodplain	HAND (m)	Horizontal Distance (m)
All Groups	2.3682	12.320	280.68
<30K	2.5920	11.108	278.60
30K - 75K	2.4103	11.498	276.88
75K - 200K	2.2603	12.862	284.53
>200K	2.3180	14.724	280.40

we should expect the lag time to be much larger.

We found quite an inconsistent change of performance between individual models and the composite model (Table I): at Dead Run, the performance of the composite model in both peak stage and lag time is better than individual models, while at Baisman Run, the composite model has poorer performance on peak stage but better performance on lag time than individual models. Since the two response variables are in different units and scales, we should assign different weights to each in the model's objective function. As the composite model structure has the potential to improve prediction accuracy than individual models at one site, we should develop a method to determine the appropriate weights for these variables for the future research. Lastly, it is obvious that our current predictors are not sufficient to predict flood timing and quantity. We should consider adding additional predictors describing rainfall locations, land cover characteristics, and storm sewer network density in the future to build a high-quality short-term flood prediction model.

B. Socioeconomic Disparity in Flood Risk

The linear regression analysis suggested that there is a relationship between the average HAND and the average price per square foot by block group, and results from the segmented regression suggested that this model might be useful in predicting property value at the block group level. The multiple linear regression model suggested that HAND and horizontal distance together are useful in predicting the average price per square foot, however some correlation between HAND and horizontal distance was observed, suggesting that a simpler model may suffice. The weighted average population in a flood plain, HAND and horizontal distances were different among different income groups (Table II).The tests for differences in population in floodplains by race yielded p-values smaller than 0.05, thus we concluded these population means are indeed different. At this time, results indicated that Whites are the most at risk. Results from the Tukey HSD test on the variable pairs indicate that there are significant differences between whites and non-whites, between black/African-Americans and non-black/African-Americans, and between those who make over 200k and the other available income brackets, but not between the lower three brackets. It is important to note however, that the populations for Asian, Pacific Islander, Native American and mixed groups were counted as very small, and there were a number of block groups that were not able to be studied due to the absence of some data. These results might warrant further study to fill in missing data to properly assess which populations are truly more at risk compared to others.

Creating maps of the features revealed that high percentages of black/African American populations seem to correspond to low HAND values (Fig. 4 and Fig. 5), a relationship that is obscured when simply plotting these values against each other, but it is important to note that some census block groups are not represented in either dataset. However, there might still be spatial patterns in flood locations with respect to socioeconomic factors that would be worth examining if these maps were overlaid with the flood prediction.

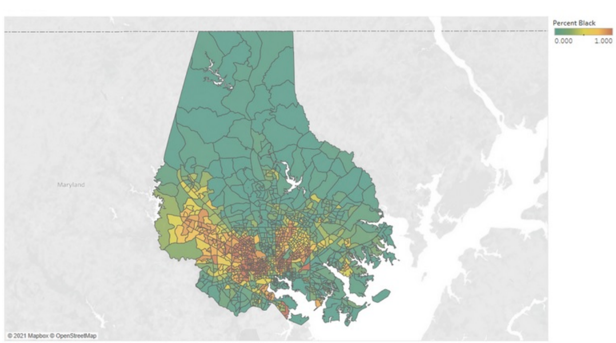

Fig. 4. Percentages of the Population that is Black/African American, by Census Block Group.

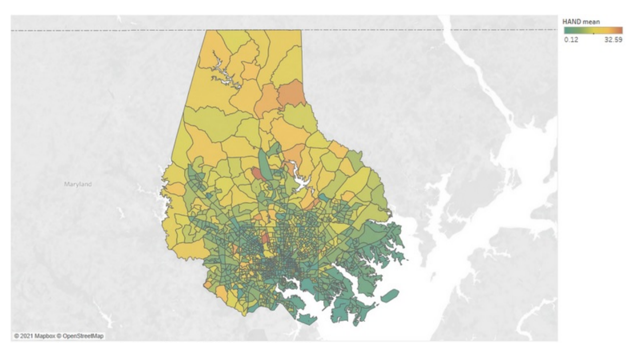

Fig. 5. The Mean Height Above Nearest Drainage, by Census Block Group.

IV. Conclusion

We implement the MLP model structure to predict 1) flood peak stage and 2) lag time between peak rainfall and stage collectively and separately at Dead Run and Baisman Run watershed in Baltimore. Using aggregated rainfall data over the entire watersheds, our model has consistently high performance on peak stages at both watersheds (best R2 over 0.8), but the performance on lag time is poor using either composite or individual models. Detailed information on precipitation centers and surface roughness, which are used in physical-based hydrological models, are missing in our model at present. In the future study, we will consider each pixel's radar rainfall intensity to counter the dynamic distance between the center of rainfall events and improve the lag-time prediction accuracy so that the model can provide high-quality flood forecasting for emergency management agencies. The socioeconomic disparity study revealed that there are connections between socioeconomic status and risk of flooding, though the severity is yet unknown. When our flood forecasting model has improved performance, we could simulate the extreme rainfall and flood events and perform the frequency analysis to determine the risk of flooding at households level resolution. There also seems to be a measurable relationship between stream proximity and property value, which underscores the hypothesis that cheaper housing, and thus those with lower income, are more at risk of being flooded in the Baltimore area. In addition, it would be beneficial to develop a metric of predicted flood damage to add to the socioeconomic analysis and peak stage prediction to better advise homeowners and city planners.

Acknowledgment

We thank US Census Bureau, USGS, and Hydrometerology Group in Princeton for sharing the data publicly and make this project possible. The high-resolution land cover data from the Chesapeake Conservancy enables us to better understand in-watershed land cover distribution and evaluate household level flood risk assessment. We also want to thank Dr. Learmonth for his generous guidance and assistance for this project. Our data and code are save at Deepnote server https://deepnote.com/project/Urban-Flood-Deep-Learning-TwLtbe1RTm-PGNxTckKEYw/%2FModelBuilding_DRKR.ipynb for the public access to view.

References

[1] K. E. Trenberth, "Changes in precipitation with climate change," *Climate Research*, vol. 47, no. 1-2, pp. 123–138, 2011.
[2] M. Campolo, P. Andreussi, and A. Soldati, "River flood forecasting with a neural network model," *Water resources research*, vol. 35, no. 4, pp. 1191–1197, 1999.
[3] M.-H. Hsu, S.-H. Lin, J.-C. Fu, S.-F. Chung, and A. S. Chen, "Longitudinal stage profiles forecasting in rivers for flash floods," *Journal of hydrology*, vol. 388, no. 3-4, pp. 426–437, 2010.
[4] K. W. Chau, C. Wu, and Y. S. Li, "Comparison of several flood forecasting models in yangtze river," *Journal of Hydrologic Engineering*, vol. 10, no. 6, pp. 485–491, 2005.
[5] S. H. Elsafi, "Artificial neural networks (anns) for flood forecasting at dongola station in the river nile, sudan," *Alexandria Engineering Journal*, vol. 53, no. 3, pp. 655–662, 2014.
[6] X. Zheng, D. R. Maidment, D. G. Tarboton, Y. Y. Liu, and P. Passalacqua, "Geoflood: Large-scale flood inundation mapping based on high-resolution terrain analysis," *Water Resources Research*, vol. 54, no. 12, pp. 10–013, 2018.
[7] D. Wagenaar, A. Curran, M. Balbi, A. Bhardwaj, R. Soden, E. Hartato, G. Mestav Sarica, L. Ruangpan, G. Molinario, and D. Lallemant, "Invited perspectives: How machine learning will change flood risk and impact assessment," *Natural Hazards and Earth System Sciences*, vol. 20, no. 4, pp. 1149–1161, 2020.
[8] S. Skakun, N. Kussul, A. Shelestov, and O. Kussul, "Flood hazard and flood risk assessment using a time series of satellite images: A case study in namibia," *Risk Analysis*, vol. 34, no. 8, pp. 1521–1537, 2014.
[9] A. Manandhar, A. Fischer, D. J. Bradley, M. Salehin, M. S. Islam, R. Hope, and D. A. Clifton, "Machine learning to evaluate impacts of flood protection in bangladesh, 1983–2014," *Water*, vol. 12, no. 2, p. 483, 2020.
[10] F. Anderson, G. L. Delclos, and D. Rao, "The effect of air pollutants and socioeconomic status on asthma in texas," *Journal of Geoscience and Environment Protection*, vol. 4, no. 9, pp. 39–52, 2016.
[11] L. P. Clark, D. B. Millet, and J. D. Marshall, "National patterns in environmental injustice and inequality: outdoor no 2 air pollution in the united states," *PloS one*, vol. 9, no. 4, p. e94431, 2014.
[12] W. H. Heiss, D. L. McGrew, and D. Sirmans, "Nexrad: next generation weather radar (wsr-88d)," *Microwave Journal*, vol. 33, no. 1, pp. 79–89, 1990.
[13] W. F. Krajewski, A. Kruger, S. Singh, B.-C. Seo, and J. A. Smith, "Hydro-nexrad-2: Real-time access to customized radar-rainfall for hydrologic applications," *Journal of Hydroinformatics*, vol. 15, no. 2, pp. 580–590, 2013.
[14] A. D. Nobre, L. A. Cuartas, M. Hodnett, C. D. Rennó, G. Rodrigues, A. Silveira, and S. Saleska, "Height above the nearest drainage–a hydrologically relevant new terrain model," *Journal of Hydrology*, vol. 404, no. 1-2, pp. 13–29, 2011.
[15] M. H. Hassoun *et al.*, *Fundamentals of artificial neural networks*. MIT press, 1995.