Predicting Maximum Flood Levels in Mburicao Stream for Enhanced Citizen Alert Systems

Mathias Aguilar [1] Héctor Velázquez [2] Diego Stalder [3] Andrés Wehrle [4]
Faculty of Engineering, University of Asunción
1 maguilar@fiuna.edu.py 2 hvelazquez@fiuna.edu.py 3 dstalder@ing.una.py 4 awehrle@ing.una.py

Introduction

Forecasting the rising water levels in the Mburicao stream is crucial for early flood prediction in the surrounding areas of Asunción. Due to its great impact, a warning systems has to be efficient enough to mitigate against dangers that flooding can pose.

The main objective of this study was to develop a machine learning model to predict peak water levels in Mburicao stream up to one hour in advance. In this case, this model can serve as an early warning system for citizens using rainfall data as its primary indicator for predicting the maximum water levels of the stream, hence facilitating effective disaster management in the area.

1 Datasets

Water level data from the Mburicao Stream, measured in meters at 10-minute intervals, were collected, along with rainfall data, measured in millimeters at the same frequency, from three stations: Aeropuerto Internacional Silvio Petirossi (AISP), Secretaría Nacional de Deporte (SND) y San Ignacio de Loyola (SIL).

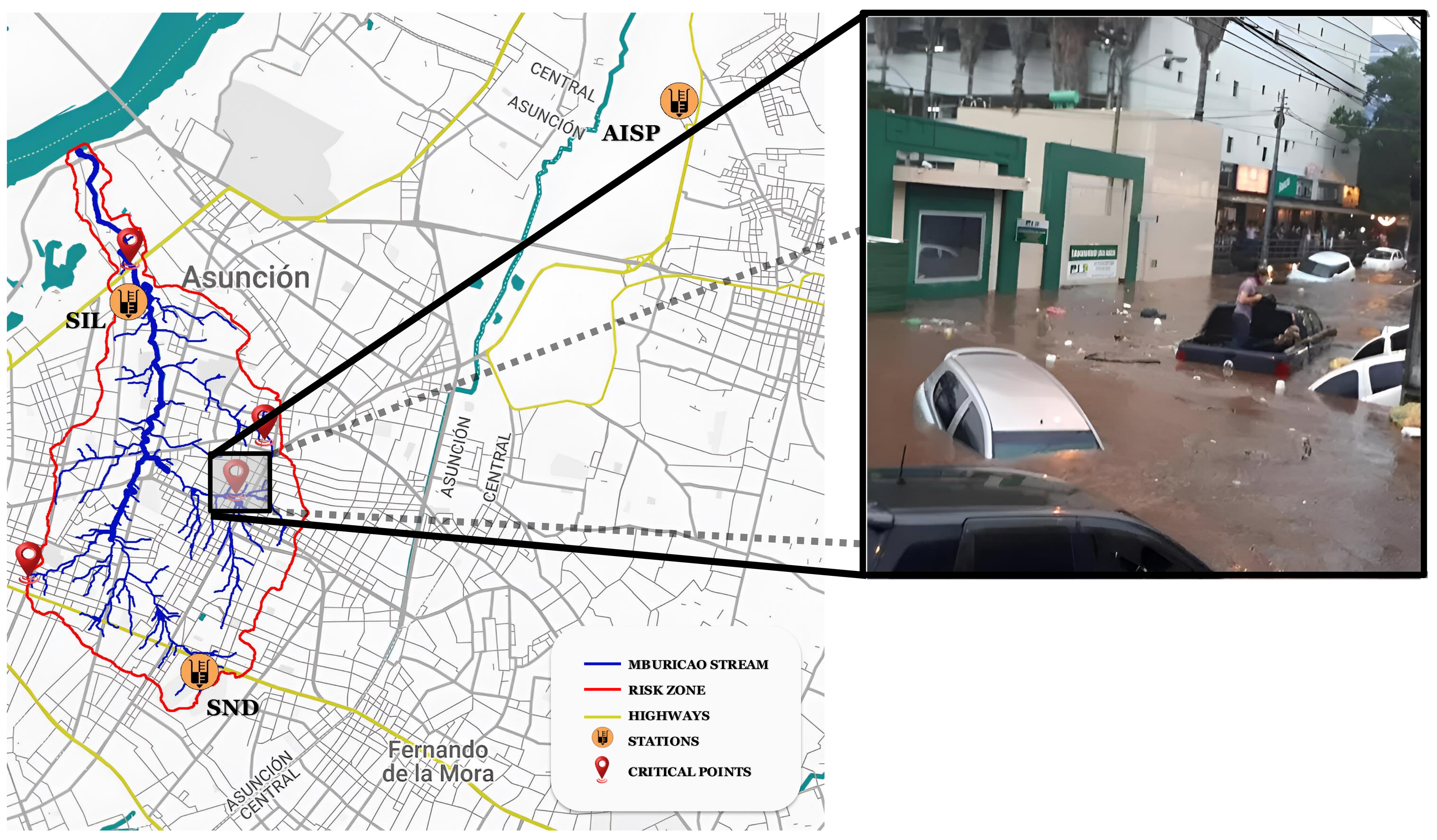

Figure 1: Map with Stations and Critical Points.

Feature/Station	City	Records Since	Records Number	Missing Values
Mburicao	Asunción	2021-05-24	50519	0
AISP	Luque	2015-01-01	316495	551
SND	Asunción	2015-05-06	376257	0
SIL	Asunción	2021-06-12	47822	0

Table 1: Datasets features by station.

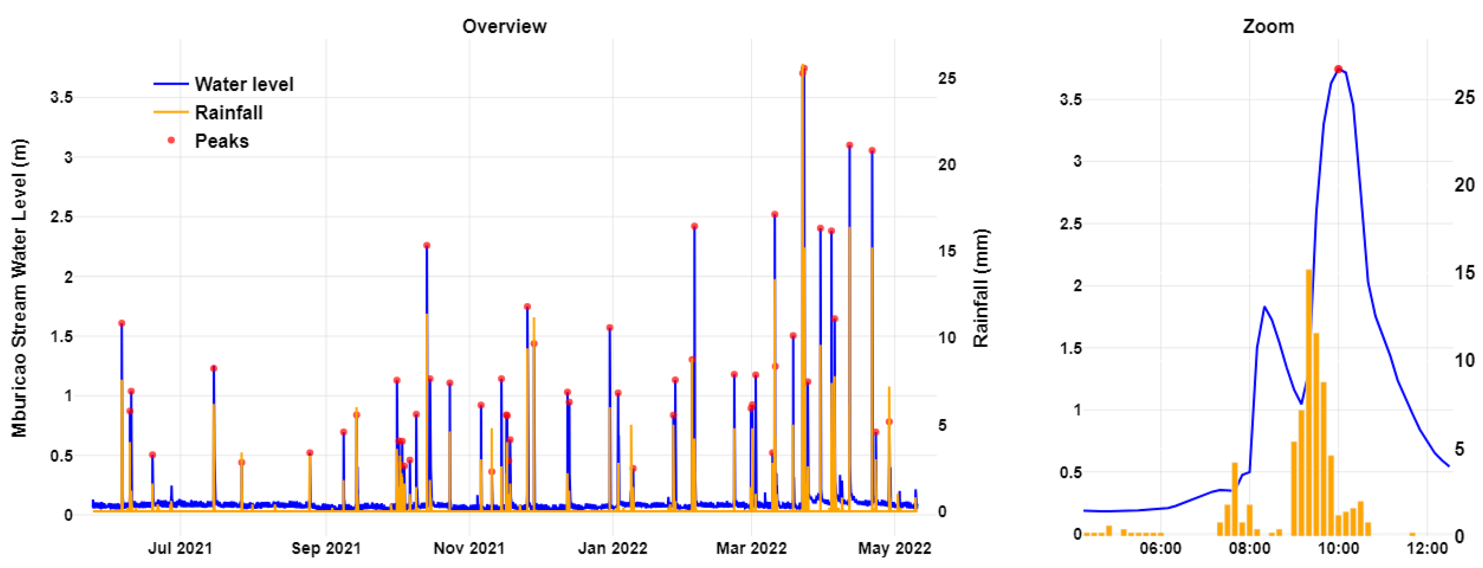

Figure 2: Line graphs of Mburicao stream water levels, highlighting peak levels and bar chart of Rainfall. Preprocessed dataset..

2 Experiments & Methodology

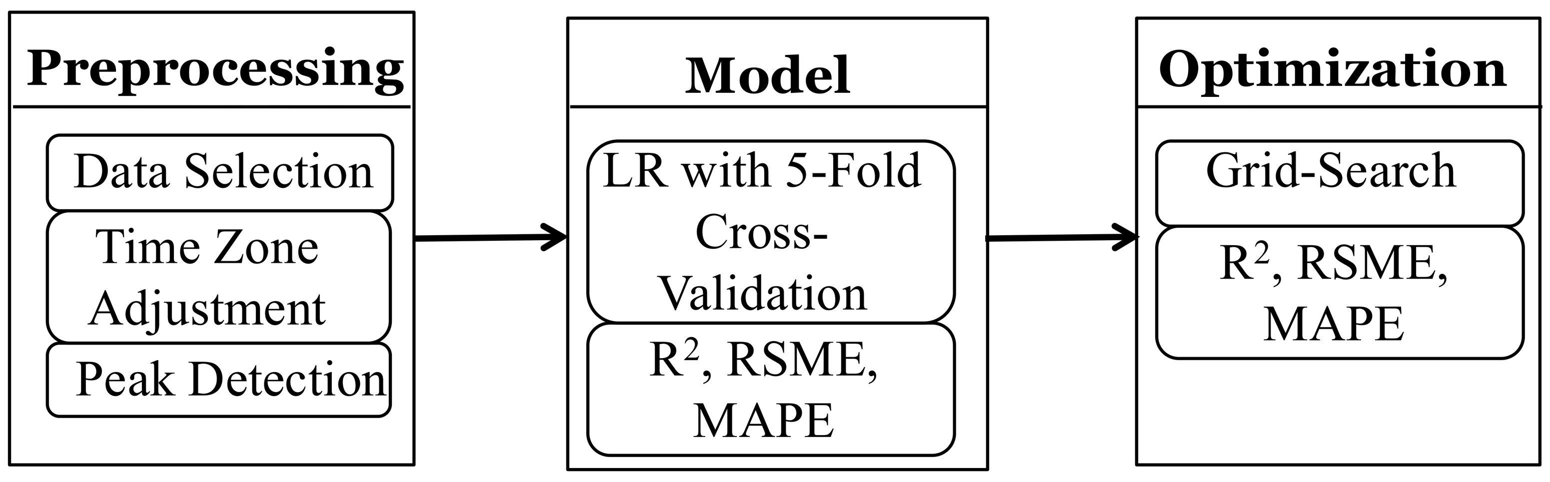

Figure 3: General flow experiment.

The experiment aimed to predict peak water levels of the Mburicao stream, identifying 54 such peaks. Events were triggered by water levels of 0.3 meters or higher. Rainfall data from part of SND and SIL stations, including up to 80 minutes prior and cumulative values, was used as input.

A linear regression model with 5-Fold Cross-Validation (k=5) was employed, using R^2, RMSE, and MAPE as metrics. Three observation windows with varying lags (W=5, L=3; W=4, L=4; W=3, L=5) were tested.

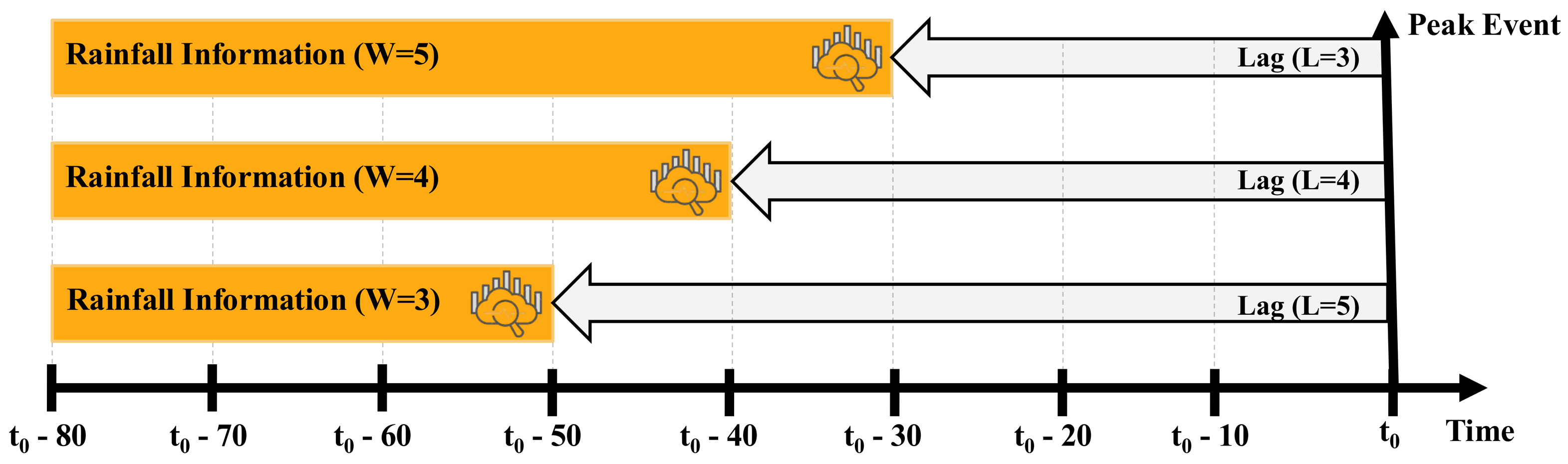

Figure 4: Visual representation of the model's predictive scope.

3 Results

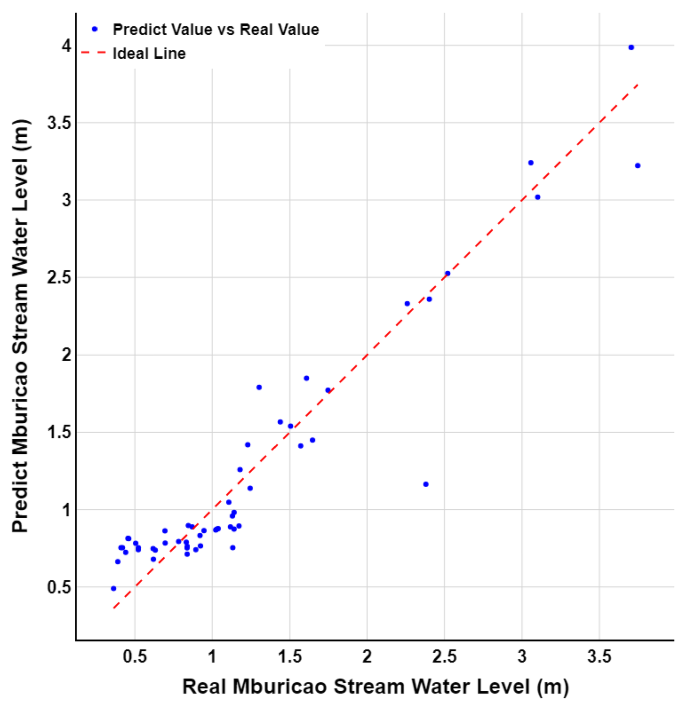

Figure 4: Scatter plot of actual vs. predicted values for W=5.

W	L	R²	RMSE	MAPE
5	3	0.8196	0.2648	22.2461
4	4	0.7758	0.2825	23.3058
3	3	0.4221	0.4477	31.9847

Table 2: Metric averages degradation of LR 5-Fold model.

W	L	Model	R²	RMSE	MAPE
3	5	LR	0.8662	0.3710	28.87
5	3	SVM	0.6744	0.5787	35.36
4	4	XGB	0.8481	0.4018	26.07

Table 3: Metric averages.Grid-Search.

The Grid-Search LR model uses rainfall data from 30-60 minutes prior. The SVM model (C=1, 0.1 tolerance) uses data from 30-50 minutes prior. The XGBoost model (0.2 learning rate, 200 estimators) uses data from 40-80 minutes prior, plus cumulative rainfall from 40 and 50 minutes.

2 Conclusions || Discussion

The model has the potential to forecast peaks 30 to 40 minutes in advance with relative efficiency, as indicated by the performance metrics. There is also potential for improvement by incorporating additional input data or optimizing hyperparameters through Grid-Search optimization.

References

1 Donald E. Knuth. *The TeXbook*. Addison-Wesley Professional, 1986.

2 Donald E. Knuth and Michael F. Plass. Breaking paragraphs into lines. *Software: Practice and Experience*, 11(11):1119–1184, 1981.