Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 12, n. 1, 2026.

Trabalho apresentado no XLIV CNMAC, Fundação Getulio Vargas - Rio de Janeiro - RJ, 2025.

Classifying Streets Susceptible to Flooding Using their Attributes and Network Metrics

Giovanni G. Soares[1] Rogério C. O. Junior[2]
INPE, São José dos Campos, São Paulo
Luiz M. Carvalho[3]
FGV, Rio de Janeiro, Rio de Janeiro
Leonardo B. L. Santos[4]
Cemaden, São José dos Campos, São Paulo

With the increasing frequency of extreme weather events due to climate change, understanding the possible causes and correlations of flooding has become essential for effective disaster preparedness and mitigation [2]. This study focuses on São Paulo, Brazil, a megacity frequently impacted by severe flooding [3], and investigates how street network attributes and centrality metrics can classify streets susceptible to flooding. Leveraging network data retrieved via OSMnx [1], we analyze centrality metrics and street attributes to train a Multilayer Perceptron (MLP) model, providing a data-driven tool to enhance urban flood risk assessment and inform mitigation strategies.

First, we retrieve the network data and attributes from the central region of São Paulo with coordinates (lat: -23.54, long: -46.63) and spanning an area of $6 \times 6km^2$. The data is represented as a graph, with nodes as intersections/dead-ends and edges as streets. Using street length as a weight, we compute edge betweenness. To obtain vertex-based metrics for streets, we transform the network into its line graph by turning edges into vertices and vice-versa. This transformation enables the calculation of key centrality measures, which are then used as features.

The features used to train the network fall into two categories: street attributes and graph metrics. Street attributes, obtained via OSMnx and Google Maps API, include characteristics such as highway type (indicating the importance of the road within the network), number of lanes, whether the street is one-way, street length, and street inclination (grade). Graph metrics, calculated using NetworkX, are derived from the line graph of the street network and include centrality measures such as Mean Shortest Path Length (MSPL), Closeness Centrality, Eigenvector Centrality, Edge Betweenness Centrality, Degree Centrality, Subgraph Centrality, and PageRank.

The dataset initially comprises 1,165 flood occurrences across the entirety of São Paulo. For this study, we focus on the central region, where 234 occurrences were recorded, affecting 120 streets (approximately 1.7% of the streets in the selected region), illustrated in Figure 1a. To address the highly imbalanced nature of the data, we randomly sampled 120 non-flooded streets, resulting in a balanced dataset with a 50% split between flooded and non-flooded streets. This approach ensures a more robust analysis and model training.

The MLP architecture consists of an input layer with 12 neurons, two hidden layers with 24 and 48 neurons, respectively, and ReLU activation functions applied after both hidden layers, leading to an output layer with two neurons. The network, illustrated in Figure 1b, is trained to classify flooded streets at any given time. The data is normalized and divided into 70% for training and validation and 30% for testing.

[1]giovanni.soares@inpe.br
[2]rogerio.manduca.jr@gmail.com
[3]luiz.fagundes@fgv.br
[4]santoslbl@gmail.com

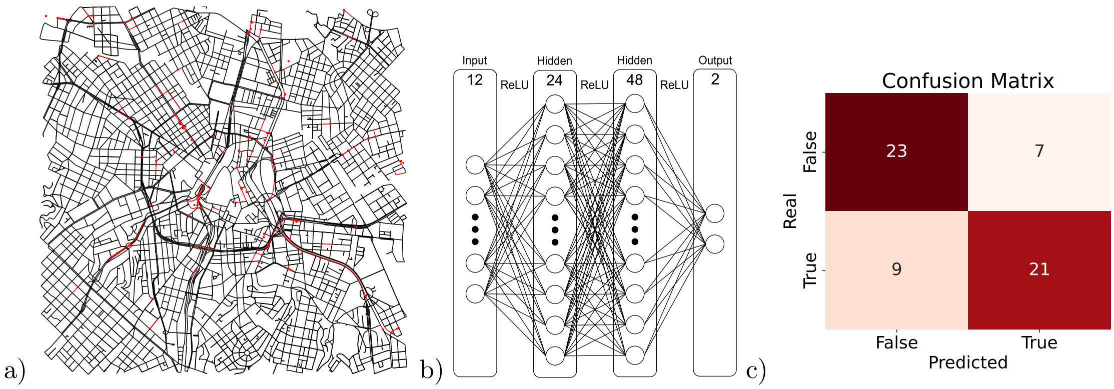

Figure 1: a) represents the street network integrated with the flood occurrence data set. b) is a graphical representation of the MLP network, with the size of each layer. c) shows the confusion matrix for the test after training, with a test accuracy of 73%. Source: Authors.

Our model achieves 73% accuracy, demonstrating the potential of street network metrics for the classification. Figure 1c brings a confusion matrix with only 7 false positives and 9 false negatives. These results highlight the potential of network-based metrics for classifying flood-prone streets and suggest that incorporating additional data, such as rainfall pattern, could further improve model performance.

Acknowledgements

We would like to thank the funding of projects 446053/2023-6, 140196/2022-6, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) PRINT project 88887.900891/2023-00.

References

[1] G. Boeing. “OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks”. In: **Computers, environment and urban systems** 65 (2017), pp. 126–139. DOI: `10.1016/j.compenvurbsys.2017.05.004`.

[2] J. S. Duque, R. Santos, J. Arteaga, R. S. Oyarzabal, and L. B. L. Santos. “Nonlinear hydrological time series modeling to forecast river level dynamics in the Rio Negro Uruguay basin”. In: **Chaos: An Interdisciplinary Journal of Nonlinear Science** 34.5 (2024). DOI: `10.1063/5.0201784`.

[3] L. R. Tomás, G. G. Soares, A. A. S. Jorge, J. F. M., V. L. S. Freitas, and L. B. L. Santos. “Flood risk map from hydrological and mobility data: A case study in São Paulo (Brazil)”. In: **Transactions in GIS** 26.5 (June 2022), pp. 2341–2365. DOI: `10.1111/tgis.12962`.