# Comparative Analysis of DEM Products for Flood Susceptibility Mapping in Urban Areas using Machine Learning

**Marcos Roberto Benso[1], Maria Clara Fava[2], Anaí Floriano Vasconcellos[3], Marina Batalini[4], Beliana Cavalcante Sawada de Carvalho[5], Alan Vaz Lopes[6], Maria Elisa Leite Costa[7], Thiago Souza Biscaro[8], Javier Tomasella[9]**
Luiz de Queiroz College of Agriculture, University of São Paulo, Piracicaba, 50200, Brazil[1]
E-mail: marcosbenso@usp.br
Federal University of São Carlos, São Carlos, São Paulo, Brazil[2]
E-mail: mcfava@ufscar.br
Federal University of São Carlos, São Carlos, São Paulo, Brazil[3]
E-mail: anai.vasconcelos@ufscar.br
Federal University of Itajubá, Itajubá, Minas Gerais, Brazil[4]
E-mail: marinamacedo@unifei.edu.br
National Institute for Space Research, São José dos Campos, São Paulo, Brazil[5]
E-mail: beliana.sawada@unesp.br
National Water and Basic Sanitation Agency, Brasília, Distrito Federal, Brazil[6]
E-mail: alanvazlopes@gmail.com
Ministry of Cities, Brasília, Distrito Federal, Brazil[7]
E-mail: elisa.costa@cidades.gov.br
National Institute for Space Research, Cachoeira Paulista, São Paulo, Brazil[8]
E-mail: thiago.biscaro@inpe.br
National Institute for Space Research, Cachoeira Paulista, São Paulo, Brazil[9]
E-mail: javier.tomasella@inpe.br

**ABSTRACT**

Data-driven flood susceptibility mapping derived from flood inventories help to evaluate the relationship between influencing factors and occurrence of urban floods. Despite recent efforts, the role of different digital elevation models has been poorly described in the literature, which have practical applications that are relevant for regions that are poorly covered by high resolution digital elevation models. This study has the objective of analysing the impact of different digital elevation models on flood susceptibility mapping using random forest. The main methodological challenge is that flood inventories usually only have positive points (flooded), but machine learning models require also negative points (non-flooded). This issue was addressed using a positive and unlabeled learning method in two steps. The first step is aimed to identify a set of relatively reliable negative (RRN) from a randomly generated set of points within the area, and the second step aims to create flood maps using the flood inventory and RRN dataset. The best model results were achieved with LiDAR 5m (Kappa 0.99); AlosPalsar, LiDAR 30m, ANADEM, and SRTM compose a middle tier (Kappa 0.92); LiDAR 10 m presents a lower performance suggesting the impact of noise in the aggregation process (Kappa 0.89); and ASTER presents the worst performance (Kappa 0.75). Considering the average of the machine learning models, the most important flood conditioning factors were Height Above Nearest Drainage (HAND) and Topographic Position Index (TPI) with windows of 35 and 25. The results suggest that the most common globally available digital elevation models can produce reliable flood susceptibility maps and that the PUL applied in this study can be successfully applied to other flood susceptible urban areas.

**KEYWORDS:** random forest, positive and unlabeled learning, variable importance

# 1 INTRODUCTION

The current state of climate change has imposed great challenge for urban water management. A warmer atmosphere can hold more water, hence the increase in the frequency and occurrence of severe heavy rainfall, especially in the tropics. From a practice point of view, urban planners and decision-makers have the hard task to continuously update the flood susceptibility maps and monitor urban areas with higher risk of floods. The flood risk mapping requires in upfront the correct description of the terrain with high resolution topographic data and digital elevation models (Tehrany et al., 2019).

Terrain analysis derived from digital elevation models (DEMs) provides a fundamental basis for identifying areas that are more likely to experience urban flooding. However, flood impacts are not determined by topography alone; exposure of human lives and assets plays a central role in shaping flood risk (Rentchler et al., 2022). Flood modeling supports the characterization of flood propagation beyond the river channel, yet flow measurements are typically confined to the channel itself, which increases uncertainty in representing flood dynamics once water spreads across the floodplain (Huang & Qin, 2014). Flood inventories have therefore been widely used to document areas affected by flooding, particularly those with the greatest impacts on the population. Although these datasets depend on civil defense reporting and response capacity, they offer significant potential for mapping flood-susceptible areas and complementing model-based approaches.

Data-driven methods can improve the flood map susceptibility with the use of flood inventories. Several studies have demonstrated the potential of machine learning algorithms for binary classification such as Random Forest, Support Vector Machine (Mukomberanwa & Madamombe, 2025), and Artificial Neural Networks (Li et al., 2022). One of the main bottlenecks of training a classification algorithm is that flood inventories only report the flood occurrence points, which leaves the non-occurrence points not reported. For binary classification, the model requires both negative and positive examples to be trained. Another question is whether the digital elevation model plays a significant role in the model training. The terrain characterization has been shown to be of great relevance for flood modeling, however, does play the same role in a1 data-driven model?

In the context of data-driven models, this study aims to assess the effect of different digital elevation models on flood susceptibility mapping using positive and unlabeled (PUB) methods for random forest. A case study is built in the Aricanduva Catchment, São Paulo, Brazil, by comparing the predictive performance of Random Forest (RF) models.

# 2 METHODOLOGY

The model framework was implemented in R environment (R Core Team, 2024) using the packages terra with methods for working with raster data and spatial data analysis (Hijmans, 2025) and sf for working with simple features providing a standardized way to encode and analyze spatial vector data (Pebesma & Bivand, 2023). The processing of terrain metrics was also performed using Whitebox tools front end interface in R (Lindsay, 2016; Wu & Brown, 2022). The machine learning models were implemented using the package caret (Classification and Regression Training) with provides functions for training and plotting classification and regression models (Kuhn, 2008). The random forest was implemented using the package randomForest (Liaw & Wiener, 2002).

## 2.1 Case study and datasets

The present study was conducted in the Aricanduva Catchment, São Paulo Metropolitan Region, Brazil, which is predominantly urban (ca. 86% according to MapBiomas). The study area covers 103 km², characterized with dense urban development and poor natural land used. The average altitude above the sea level is 750 meters.

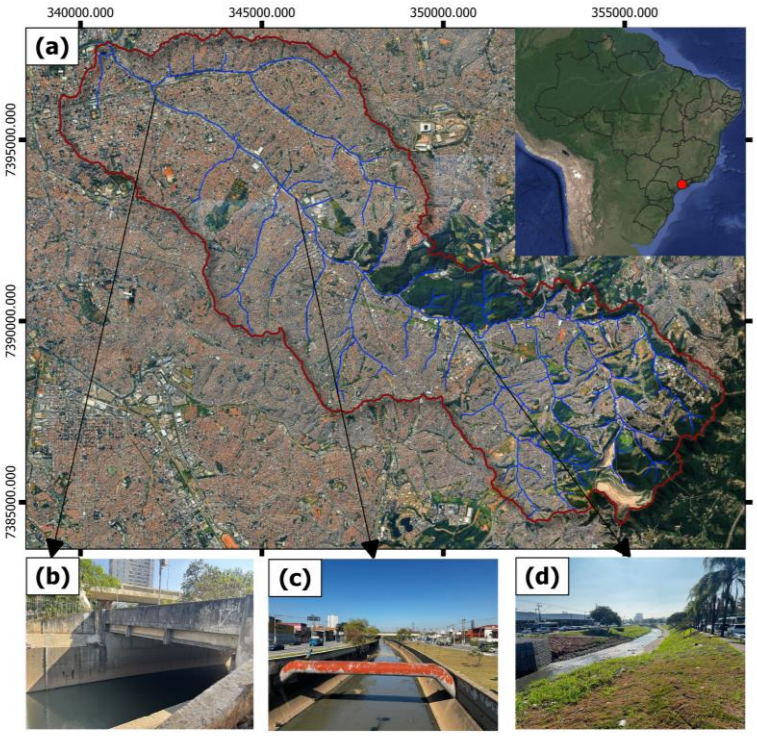


Figure 1: (a) Delineation of the Aricanduva Catchment area with highlights of (b) channel nearby the catchment outlet in the Tietê River, channel (c) in the middle part and (d) in the upper part of the catchment.

The flood inventory database was collected from São Paulo Municipal Geographic Data (Web-1). The flood inventory is collected by the civil defence since 2013 considering the official definition of drainage system overflow and floods. The drainage system overflow could be related to the fluvial processing during floods and inefficiencies of the drainage system during heavy rainfall. In the time from of 2013 to now, 593 flood occurrences were registered in the Aricanduva catchment area with most of the events being concentrated the month of February (Figure 2). The wet period is from October to March, where most occurrences were registered.

This study evaluates a set of widely used and freely available Digital Elevation Model (DEM) products. ANADEM, SRTM, NASADEM, and ALOS PALSAR. alongside LiDAR-derived DEMs generated at different spatial resolutions (5 m, 10 m, and 30 m). These datasets span a range of sensor technologies, processing methodologies, and levels of vertical and horizontal accuracy, offering a representative basis for comparison in flood modeling applications.

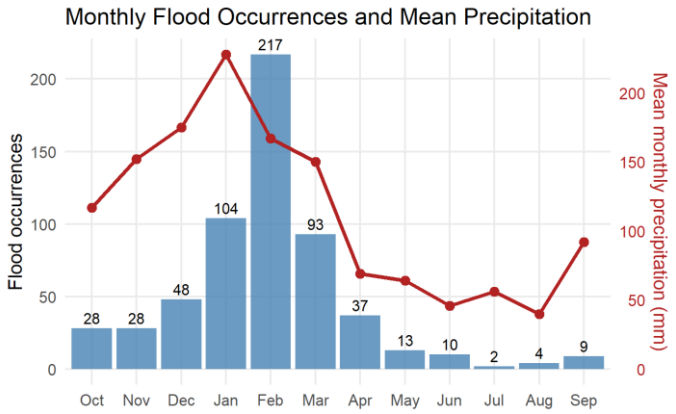


Figure 2: Annual distribution of monthly count of flood occurrences and historic mean precipitation in the Aricanduva Catchment

## 2.2 Flood conditioning factors

The conditioning factors considered in this study include a set of topographic and hydrological variables derived from the digital elevation model. Slope (°) expresses the terrain inclination in degrees, while Aspect (°) indicates the compass direction or azimuth that the terrain surface faces; both are continuous variables.

To characterize terrain morphology, several indices were used. The Terrain Ruggedness Index (TRI) (°) quantifies terrain roughness and is continuous. The Topographic Position Index (TPI) (°) describes the relative position of a pixel in relation to its surroundings and was computed using different moving window sizes: 8 × 8 pixels (Queen case), 3 × 3 pixels (TPI 3), 15 × 15 pixels (TPI 15), 25 × 25 pixels (TPI 25), and 35 × 35 pixels (TPI 35). All TPI-derived variables are continuous. Roughness (°), representing terrain roughness measured in degrees, is also a continuous variable.

Hydrological characteristics were represented by Flow Direction (FLD) and Eight Flow Direction (D8), are discrete variables that indicates the direction of surface runoff, and HAND (m), which expresses the height above the nearest drainage channel and is treated as a continuous variable.

## 2.3 Positive and unlabeled classification of flood susceptibility

The problem of positive and unlabeled classification is common to flood inventory data where only the positive, i.e., the point of flood occurrence is known ($y = 1$). The negative data, which are the points where the flood did not occur, generally are not recorded. The methodology proposed is divided into two steps: (i) the first step is designed to obtain a set of relatively reliable negative (RRN) points using a machine learning model, and (ii) the second step has the objective of using the new dataset composed of RRN to train a new machine learning model to generate the flood susceptibility maps. The models in the two steps were trained using random forest and the hyperparameter tunning considere1d leave-one-out cross-validation (LOOCV) to maximize data usage and reduce bias associated with limited positive samples. Class probabilities $P(s = 1|x)$ were estimated for each sample, and model performance was evaluated using two-class summary statistics based on receiver operating characteristic (ROC) analysis.

Model performance was evaluated using precision, accuracy, and Cohen's Kappa. Precision was computed as the ratio of correctly predicted flooded locations to all predicted flooded locations. Accuracy represents the proportion of correctly classified samples, while Cohen's Kappa measures the agreement between predictions and observations corrected for chance.

In the step-one, a set of random points ($s$) is generated over the study area. There is a chance of each random point being a flood point ($s = 1$) and non-flood point ($s = 0$). The random points were generated 10 times the number of flood points. For training the model, an unbalanced dataset was created considering 80% of the positive dataset and 2% of the random points. The test dataset was composed only with the 20% of the positive dataset. The main goal of the first model is to create a bias towards the positive prediction. The trained model is then used to classify the rest of 98% of random points into 0 and 1 probabilistically. The probabilistic prediction over the observed flood points is compared with the observed positive points, that is, the model predicts $p(x) = P(y = 1|x)$ for the cases we know the flood points. To choose the RNN points, a threshold representing the minimum probability required for a point to be considered flooded is given by equation 1. The RRN are given by the $P(s = 1|x) \leq c$, that is, all predictions made over the random points that were classified with a probability lower than the threshold.

$$c = \min_{x \in P} P(y = 1|x) \quad (1)$$

In the step-two, a new dataset is created to train random forest models to perform flood susceptibility mapping. The division of the observed positive data (flood points) is maintained the same as the step-one. The same 80% of positive labels is used for training and 20% used for testing the model in step-one was used for the models in step-two. The difference is that the RRN points were now randomly sample in 80% for training and 20% for testing.

## 3 RESULTS AND DISCUSSION

A comparison of model performance for test set for step-one and step-two is presented on Table 1. The test set is the 20% portion of the dataset that was never used for training the machine learning models. When considering precision, both steps presented a high performance. Step-one presented an equivalent or higher precision than step-two. This is because we introduced a bias towards classification of positive cases in the step-one. This result shows that the models were overoptimistic toward positive classification, which indicates a good classification of relatively reliable negatives. When comparing accuracy and kappa, it is clear that step-two presented a highly superior performance than step-one when considering both classes, positive and RRN.

The reduction of LiDAR resolution impacts model performance, however, there is an especial attention that must be paid to the impact of noise that can be introduced by the aggregation method. LiDAR 5m presented the highest performance with almost perfect classification metrics and was able to classify the highest number of relatively reliable negatives. The changes in precision in the two steps of LiDAR 5m indicate that the RRN did not reduced the ability of the model of classifying positive correctly and the increase in accuracy and kappa corroborate that there was an overall improvement of the model.

Considering a balance among the performance metrics, there is a middle tier with very similar performance models LiDAR 30m, ANADEM, SRTM and ALOSPALSAR. ALOSPALSAR presents a lower precision, which is related to the ability of the model predict correctly positive points. The worst performing model was trained with ASTER.

Table 1: performances of random forest model trained with different digital elevation models

| Name | Precision | | Accuracy | | Kappa | | Nº of RRN |
|---|---|---|---|---|---|---|---|
| | Step-one | Step-two | Step-one | Step-two | Step-one | Step-two | |
| LiDAR 5 m | 0.999 | 0.994 | 0.258 | 0.996 | 0.013 | 0.993 | 167 |
| LiDAR 10 m | 0.997 | 0.868 | 0.364 | 0.955 | 0.019 | 0.896 | 46 |
| LiDAR 30 m | 0.997 | 0.907 | 0.386 | 0.962 | 0.020 | 0.920 | 49 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| ANADEM | 0.995 | 0.912 | 0.339 | 0.959 | 0.015 | 0.917 | 62 |
| SRTM | 0.994 | 0.914 | 0.360 | 0.956 | 0.015 | 0.911 | 74 |
| ASTER | 0.996 | 0.636 | 0.190 | 0.956 | 0.006 | 0.755 | 7 |
| ALOSPALSAR | 0.996 | 0.892 | 0.189 | 0.972 | 0.007 | 0.924 | 33 |

The probabilistic binary predictions, that is, the probability of a prediction being a flooded point, for all digital elevation models are shown in Figure 3. The Aricanduva catchment was classified into five susceptibility classes; very low, low, moderate, high and very high. The visual analysis show that the high and very high flood susceptibility areas are located around the main channel of the Aricanduva and around the outlet.

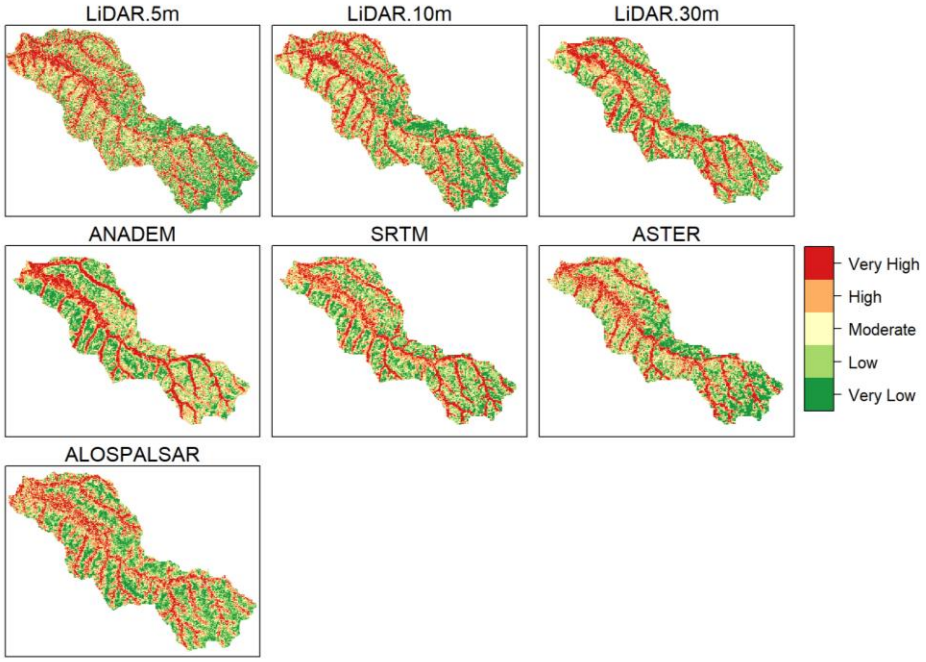


Figure 3: Flood susceptibility predictions produced by random forest model.

The variable importance that is extracted from the trained random forest in the step-two. The most important variable for all models was the height above the nearest drainage (HAND). The topographic position index (TPI) also played a significant role in model predictions. The number 35, 25 and 15 are the number of neighbouring cells, and, as observed in Table 2, the larger the number of neighbouring cells the more importance in model prediction. For all models, the D8 and flow direction (FLD) are not significatively important for model prediction and, for most models, FLD was unused for model predictions.

Table 2: Relative importance of terrain covariates (ordered by mean importance) extracted from the step-two random forest

| **Covariate** | **LiDAR 5 m** | **LiDAR 10 m** | **LiDAR 30 m** | **ANADEM** | **SRTM** | **ASTER** | **ALOS PALSAR** | **Mean** |
|---|---|---|---|---|---|---|---|---|
| HAND | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | **100.0** |
| TPI_35 | 62.3 | 45.4 | 52.7 | 58.3 | 84.3 | 89.3 | 61.5 | **64.8** |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| TPI_25 | 56.9 | 26.3 | 32.4 | 50.0 | 58.5 | 94.4 | 49.4 | **52.6** |
| TPI_15 | 42.7 | 25.1 | 28.3 | 57.4 | 59.1 | 78.9 | 41.3 | **47.5** |
| TRI | 53.8 | 31.8 | 42.5 | 35.5 | 33.9 | 55.9 | 17.8 | **38.7** |
| ROUGH | 60.7 | 29.4 | 24.7 | 32.0 | 38.8 | 57.0 | 0.0 | **34.7** |
| SLOPE | 58.2 | 24.6 | 19.2 | 31.0 | 35.8 | 53.6 | 44.7 | **38.2** |
| TPI | 45.0 | 15.0 | 40.4 | 30.8 | 42.6 | 64.0 | 9.7 | **35.4** |
| TPI_3 | 38.4 | 16.1 | 36.0 | 29.9 | 44.4 | 46.5 | 34.5 | **35.1** |
| ASPECT | 32.6 | 22.1 | 27.5 | 20.1 | 31.8 | 41.3 | 32.5 | **29.7** |
| D8 | 9.6 | 5.8 | 9.1 | 7.7 | 9.7 | 13.3 | 15.4 | **10.1** |
| FLD | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.4 | **0.3** |

## 4 CONCLUSION

This study investigated the use of different digital elevation models for flood susceptibility mapping using random forest. The machine learning models were developed in positive in unlabeled learning framework divided into two steps that is typical of flood inventory data that usually has only the positive cases (flooded points). The step-one generated a set of relatively reliable negative data points that were used the step-two to predict flood susceptibility. The results indicate that the random forest were able to correctly classify flood-prone areas, that were mostly controlled by the height above the nearest drainage and topographic position index that consider the influence of higher number of neighbouring cells.

The comparison of products indicates that LiDAR 5m, which was the lowest resolution, presented the best performance with precision, accuracy and kappa of 0.99. The reduction of horizontal resolution impacted model results, however, there was a good agreement that most used global digital models ALOS PALSAR, and SRTM and the Brazilian model ANADEM, presented high performance metrics with precision close to 0.90, accuracy to 0.96 and kappa close to 0.92. The results indicate that these models can be used for reliable flood mapping based on flood inventory in other study areas. The weakest model was produced by ASTER data with precision of 0.63, accuracy of 0.95 and kappa of 0.75.

Flood susceptibility mapping relies on historical flood records and static topographic variables and therefore does not explicitly represent temporal changes in land use, urban drainage capacity, or rainfall characteristics. Moreover, uncertainties related to input data resolution and the completeness of reported flood events may influence model performance. Future research should prioritize the integration of dynamic rainfall information, enhanced representation of urban drainage networks, and the coupling of the proposed framework with real-time hydraulic simulations to support operational flood early warning systems.

## 5 ACKNOWLEDGEMENTS

The authors acknowledge the support of the National Council for Scientific and Technological Development (CNPq) through research grants (processes 381989/2024-0 and 174232/2023-3) and funding provided by the CNPq/MCTI Call No. 15/2023 – Extreme Meteorological Events: Natural Disaster Prevention and Damage Minimization (process 446029/2023-8).

## REFERENCES

Hijmans R (2025). _terra: Spatial Data Analysis. R package version 1.8-54, <https://CRAN.R-project.org/package=terra>.

Huang, Y., & Qin, X. (2014). Uncertainty analysis for flood inundation modelling with a random floodplain roughness field. Environmental Systems Research, 3(1), 9. https://doi.org/10.1186/2193-2697-3-9.
Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. Journal of Statistical Software, 28(5), 1–26. https://doi.org/10.18637/jss.v028.i05.
Li, W., Liu, Y., Liu, Z., Gao, Z., Huang, H., & Huang, W. (2022). A Positive-Unlabeled Learning Algorithm for Urban Flood Susceptibility Modeling. Land, 11(11), 1971. https://doi.org/10.3390/land11111971
Liaw, A. and Wiener, M. (2002). Classification and Regression by randomForest. R News 2(3), 18--22.
Lindsay, J. B. (2016). Whitebox GAT: A case study in geomorphometric analysis. Computers & Geosciences, 95, 75-84. doi: http://dx.doi.org/10.1016/j.cageo.2016.07.003.
Mukomberanwa, N. T., & Madamombe, H. K. (2025). Next generation data-driven flood susceptibility modelling with spatial machine learning. Scientific African, e03082. https://doi.org/10.1016/j.sciaf.2025.e03082.
Pebesma, E., & Bivand, R. (2023). Spatial Data Science: With Applications in R. Chapman and Hall/CRC. https://doi.org/10.1201/9780429459016
R Core Team (2024). _R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
Rentschler, J., Salhab, M., & Jafino, B. A. (2022). Flood exposure and poverty in 188 countries. Nature communications, 13(1), 3527. https://doi.org/10.1038/s41467-022-30727-4.
Tehrany, M. S., Jones, S., & Shabani, F. (2019). Identifying the essential flood conditioning factors for flood prone area mapping using machine learning techniques. Catena, 175, 174-192. https://doi.org/10.1016/j.catena.2018.12.011.
Wu, Q., Brown, A. (2022). whitebox: 'WhiteboxTools' R Frontend. R package version 2.2.0 <https://CRAN.R-project.org/package=whitebox>

Web sites:
Web-1: https://metadados.geosampa.prefeitura.sp.gov.br/geonetwork/srv/por/catalog.search#/home, consulted 18 December 2025.