Results

3.1 Selection of the landslides conditional factors

To determine which factors have contributed to the quality of the model, these were evaluated by the IGR technique in the area of study. Figure 5 illustrates the results of the IGR index for the 12 factors selected in the study area, all of which are greater than 0. The findings show that the Valley Index (VD) has the most significant predictive capacity for the model. On the other hand, the Melton index have the lowest value. Other factors, which include TWI, NDGI, TPI significantly contribute to the landslide model. In contrast, the 12 factors selected (aspect, elevation, hillshade, total and planar curvature, Factors such as Slope Size, TRI, Ruggedness, NDMI, NBRI, BSI and LS Factor have a merit value equal to 0, which results in their exclusion from the modelling process. This is due to the detrimental effect of introducing noise into the model, which reduces the predictive ability of the model (Tien Bui et al., 2016).

Additionally, correlation among the 12 factors chosen in the previous stage is measured removing from the analysis those that have a lower impact and related with other factors that have greater impact on the model. Under this perspective, the Melton factor, Geomorphons, NDVI, NDWI and GNDVI are excluded from the analysis, so that seven factors are finally used to build the model: VD, TWI, NDGI, TPI, convergence index, planar curvature and EVI. In this work, two factors that have not been considered in the literature are included in the final model: the NDGI (Normalized Difference Glacier Index) and the EVI (Enhanced Vegetation Index)

3.2 Models analysis

In this study, machine learning models were implemented by using the R programming language through the mlr3 package [?], which is a complete machine learning models analysis ecosystem. Optimal values of the hyperparameters obtained through the method are shown in table 6.

3.3 Model performance and validation

In the model´s evaluation, factors including the average ROC curve among all iterations produced by the cross validation, and the respective area created under the curve AUROC, are used. AUROC values vary between 0.5 and 1, where 0.5 implies having a precision identical to a model set randomly, while 1 represents the optimal model with the maximum area under the curve. Figure 7 summarizes the ROC curves for the test set, and figure 8 shows the difference between the AUC of the training set and the test set for all models. This shows that in the training set the solution was overfitted.

Taking into consideration the hyperparameters shown in table II and using the factors that contain the most information, SVM models, logistic regression, RF and XGBoost are obtained. The AUC value average result is shown in table III using cross-validation. Figure 9 also shows a box plot that allows comparing the model´s statistical distribution with respect to the classification error. This graphic shows the values obtained for the process repeated 500 times (5-fold cross-validation with 100 repetitions), including the mean and the value distribution. This metric is obtained in the same process as the AUC. In regard to AUC (Table 7), RF obtains the highest value. However, relying only on this metric might not be the optimal strategy, since higher values of AUC do not necessarily guarantee a higher spatial accuracy of the models (Aguirre, 2013). Therefore, other additional metrics of statistical evaluation are needed, like classification error. The results obtained (AUC greater than 0.9) confirm what has been shown in the literature, in the sense that both RF and XGBoost are algorithms that perform well when working with landslide susceptibility

For the statistical analysis, table 8 summarizes the results of Friedman´s overall test considering the 500 observations of the classification error for each model. Finally, table 9 shows Nemenyi´s post-hoc test results allowing the comparison among each of the models. The Friedman statistical test showed that there are significant differences among the methods. Then, using the Nemenyi pairwise test, it can be seen, that although RF have significant statistical differences with all the rest of models.

3.4 Susceptibility Maps

After evaluating the performance of the four prediction methods, the respective landslide susceptibility maps were made. To
do so, the following steps are taken:

– Ten million points are generated in the basin polygon, evenly distributed.
– At each of these points, the values of the factors causing the landslides (VD, TWI, NDGI, TPI, EVI, Convergence Index
and planar curvature) are calculated.
– Using the machine learning models, the landslide susceptibility indexes are calculated for each point.
– The points are transformed into a georeferenced raster file.
– The values obtained in step three are reclassified into regular intervals ranging from 0 to 1, using the following labels:
very low susceptibility, low susceptibility, middle susceptibility, high susceptibility, and very high susceptibility

Figures 10-13 show the maps generated by the four models under study. As seen in the figure, the four maps indicate similar areas of susceptibility. The main difference is that both the random forest and the XGBoost show greater detail than the SVM and the logistic regression. For the calculation of thresholds, the Jenks Breaks method will be used, which is widely used in the literature, as it is based on an optimization algorithm that minimizes the within-class variance and maximizes the between-class variance. Attached is an example map representing the Random Forest model.