Supplementary Material

A comparative assessment of machine learning models for landslide susceptibility mapping in the rugged terrain of northern Pakistan

Naeem Shahzad [1], Xiaoli Ding [1] and Sawaid Abbas [1,2]

[1] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong SAR, China; naeem.shahzad@connect.polyu.hk (N.S.), xl.ding@polyu.edu.hk (X.D.), sawaid.abbas@connect.polyu.hk (S.A.).
[2] Remote Sensing, GIS and Climatic Research Lab (RSGCRL), National Center of GIS and Space Applications, university of the Punjab, Lahore, Pakistan

Contents of this file

Optimization of parameters of investigated machine learning algorithms.

Introduction

Supplementary material showing the changing trends in parametric values of corresponding machine learning algorithms to find out optimal values for training and building models. For each ML model, the tuning was performed to find optimal parameter values for accurate modeling and to avoid overfitting of the models with the discussed variables and used training data on the basis of their accuracies. Each ML model, exhibits, different parameters, and the values of these parameters effects the accuracy and reliability of these models. Considering the discussion of these dependent parameters of each model in the section 2.6 of the main manuscript, the effect of changes in these parameter values on the model, w.r.t. their accuracies in some models and the Akaike Information Criterion (AIC) in maxENT model, are presented here in Figure S1 to S5.

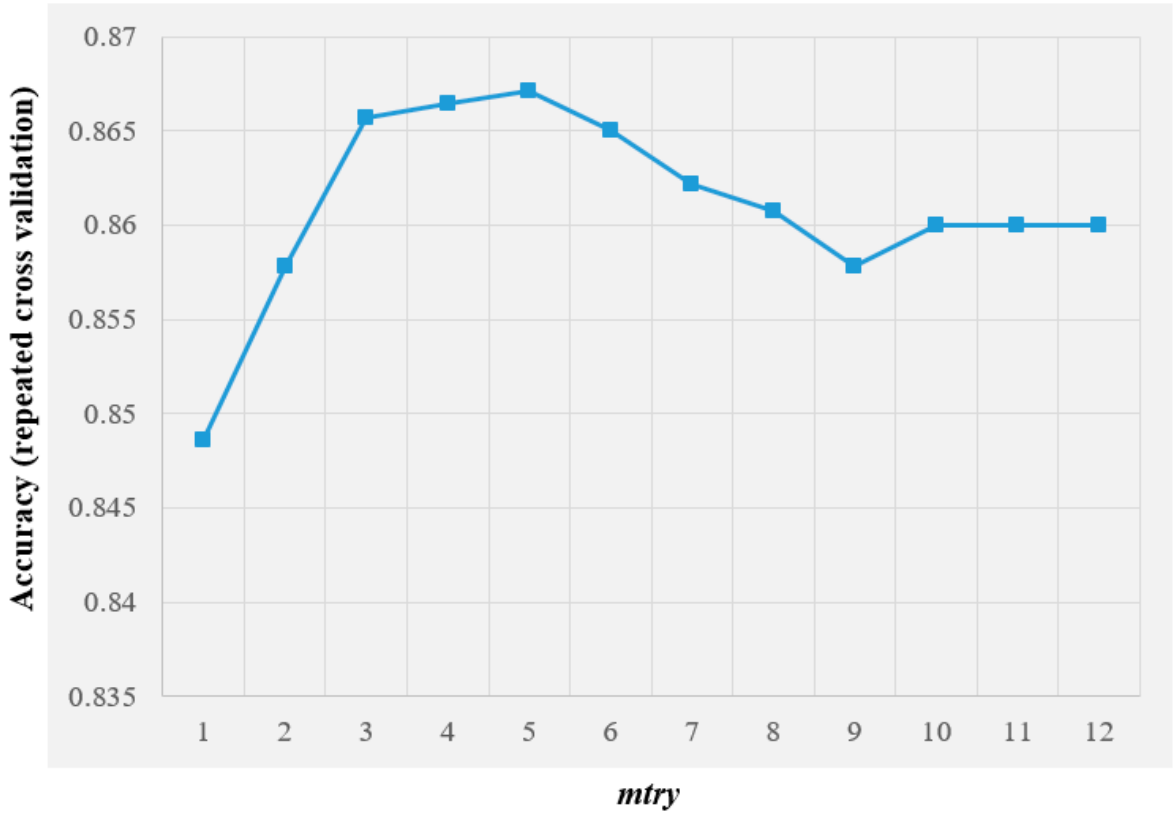


Figure S1: Changing trends in mtry against ntree=800 in Random Forest. mtry = 5 was used for training RF model.

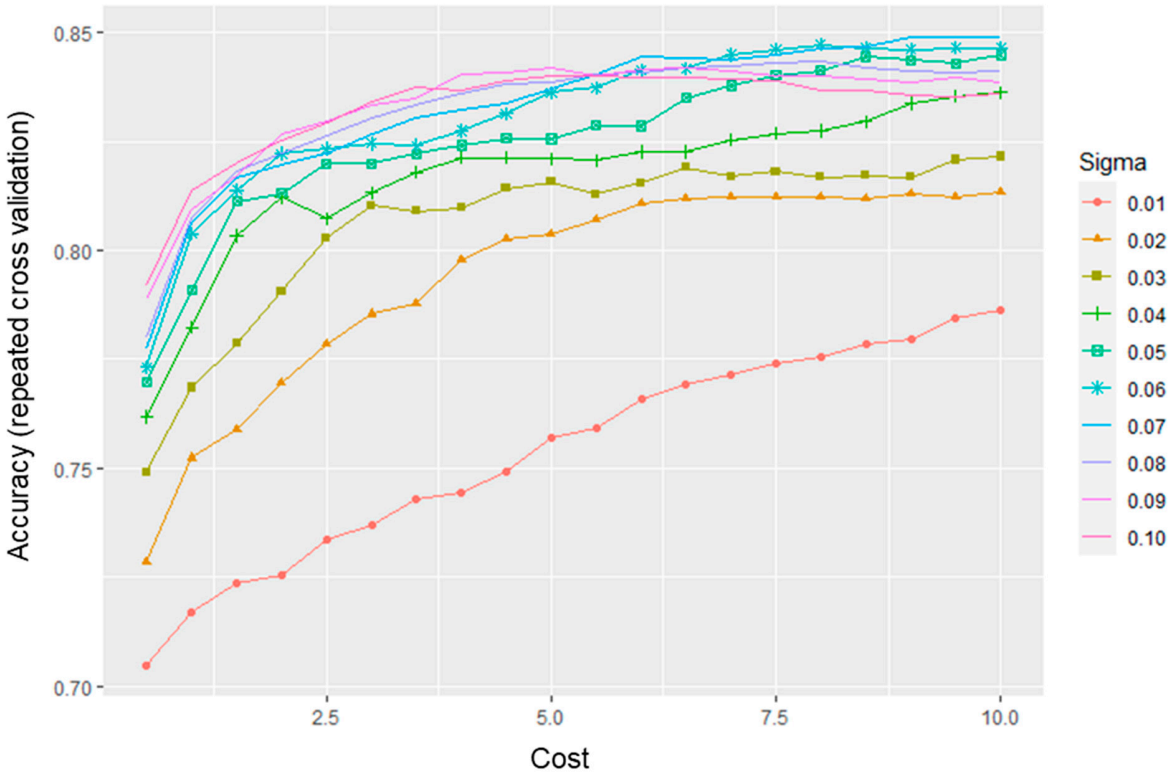


Figure S2: Optimization of cost and sigma values for SVM. C=9 and sigma =0.07 were used for training.

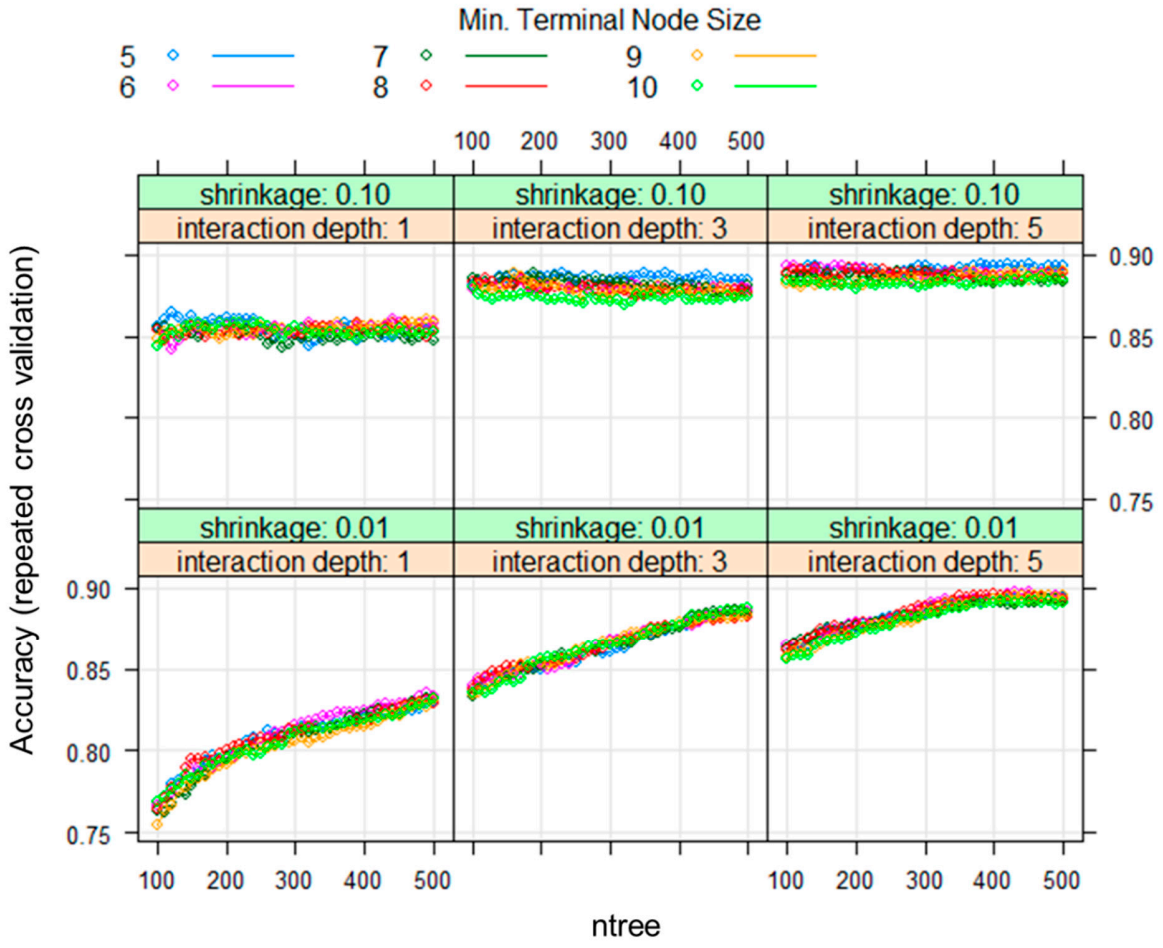


Figure S3: Changing trends in parametric values of GBM. Parametric values showing highest accuracies were used for model training. The final values used for the model training are; ntree= 460, interaction depth =5, shrinkage = 0.01 and node size of 6.

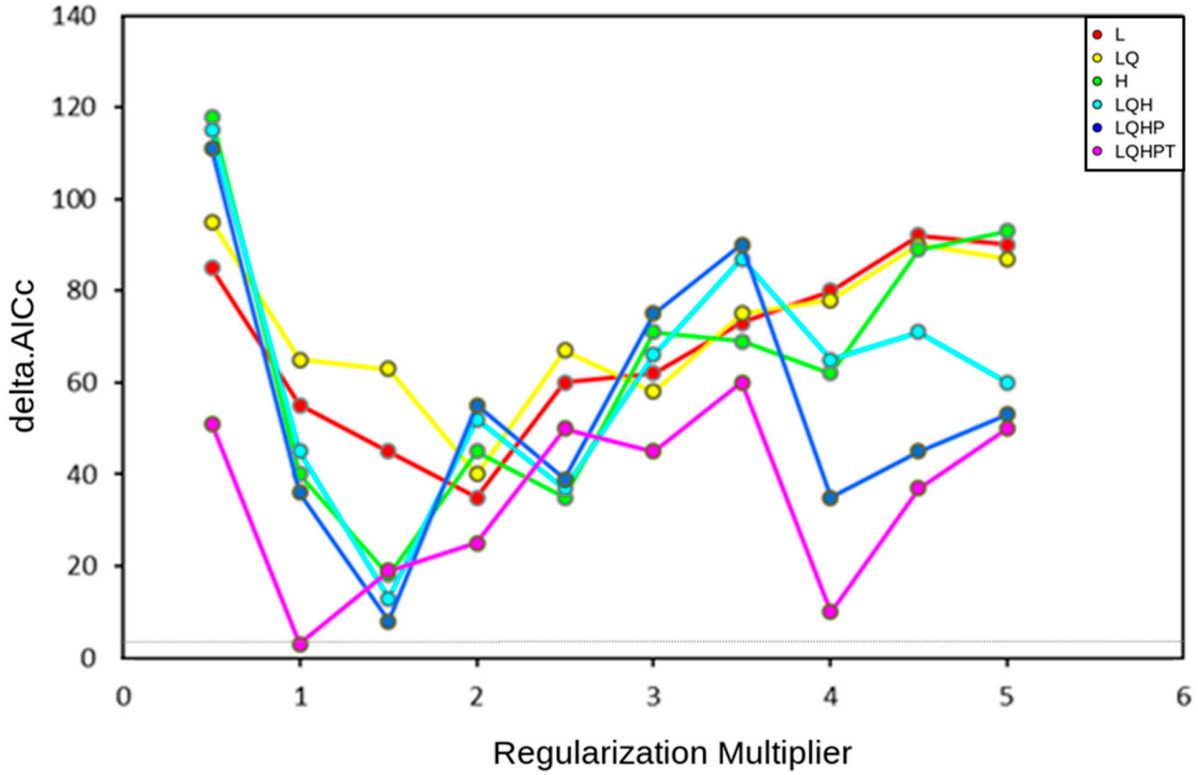


Figure S4: Changing trends in parametric value of maxENT for different feature types. Parametric values showing lowest AICs were used for model training. According to the literature, the values close to zero are considered to provide bet model fitting. Here, L is linear, Q represents quadratic, H for hinge, P for product and T represents threshold features. Regularization parameter = 1 showed the minimal AIC and used for modeling.

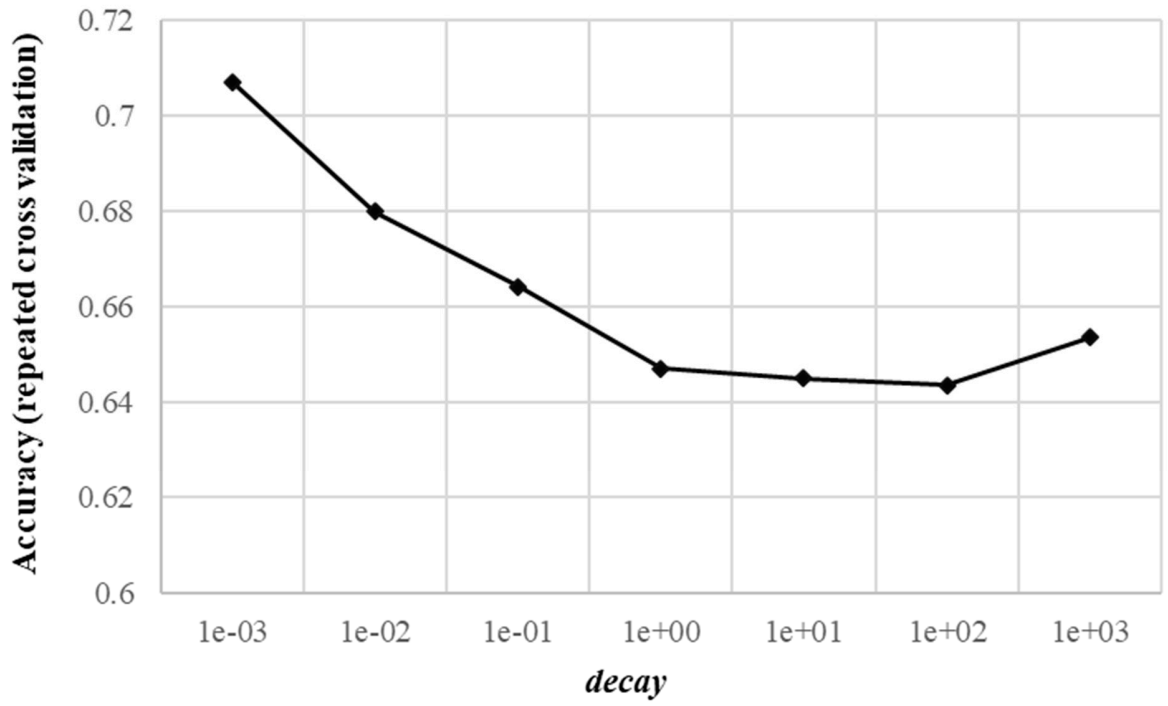


Figure S5: Changing trends in parametric value of decay for LR model. Decay = 0.001 was for modeling.