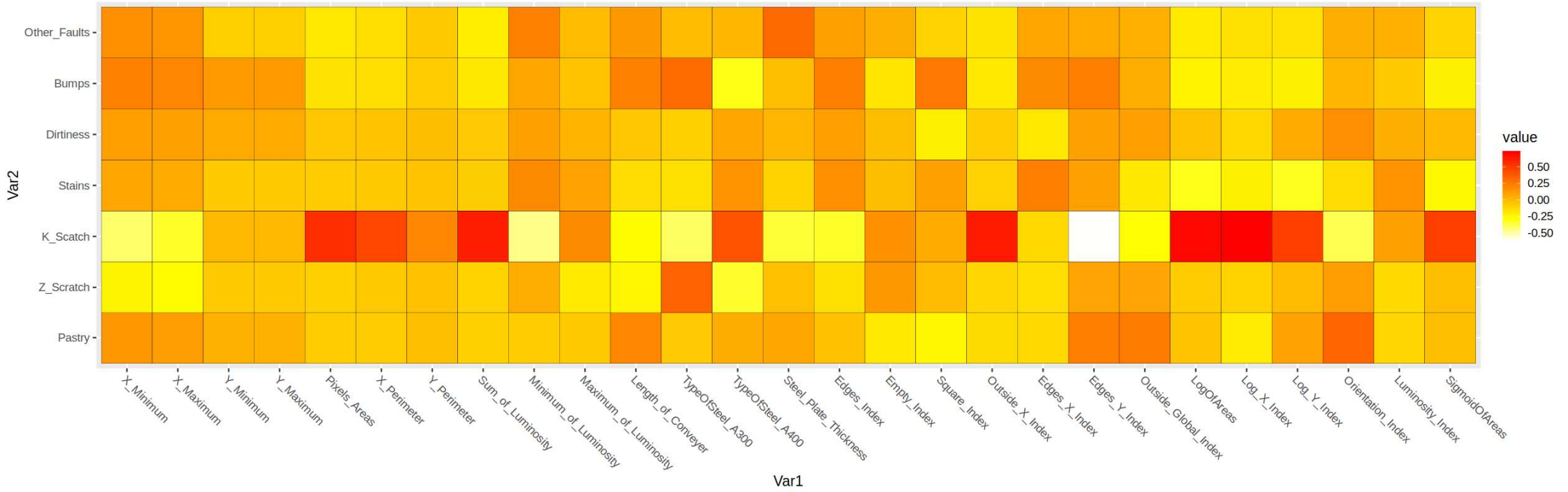
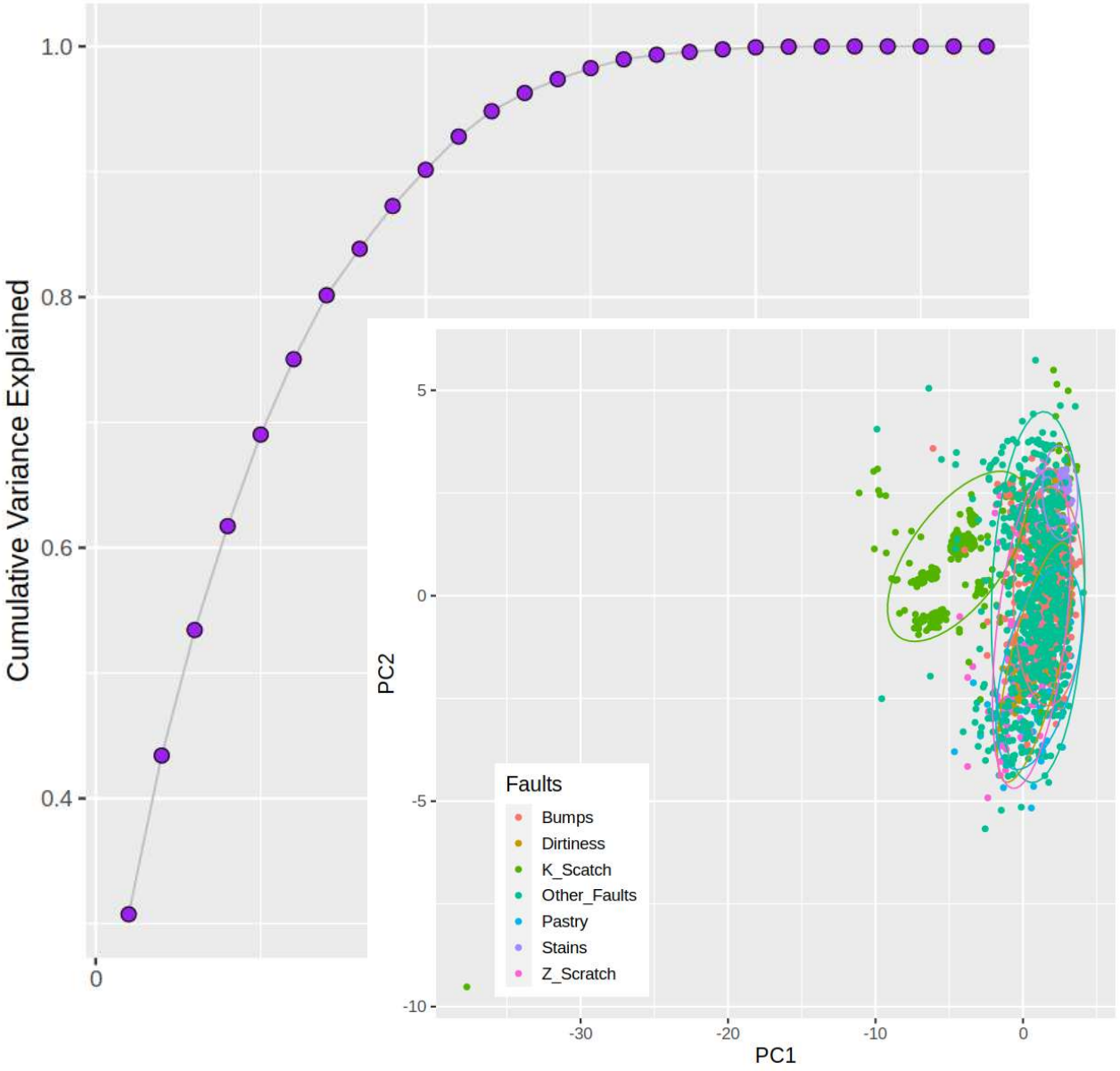
SDSC3006 Project
Classification of Steel Plate Faults



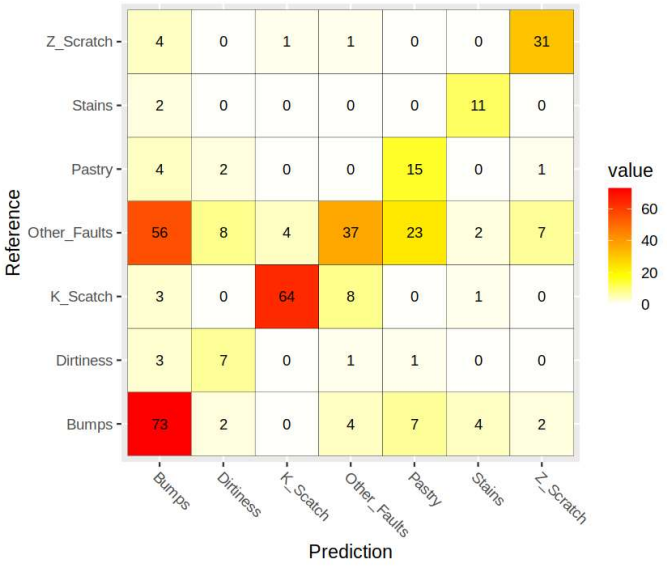
Background

- 27 features x 7 types
  - Predict faults of steel plates
- PCA
  - Reduce dimensions
- N = 12
  - 95% variance of the data
- Goal
  - Reduce noise
  - Increase accuracy
- Training set
  - 80% of the data
- Testing set
  - 20% of the data

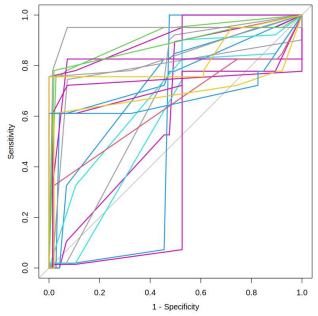


Naive Bayes Classifier

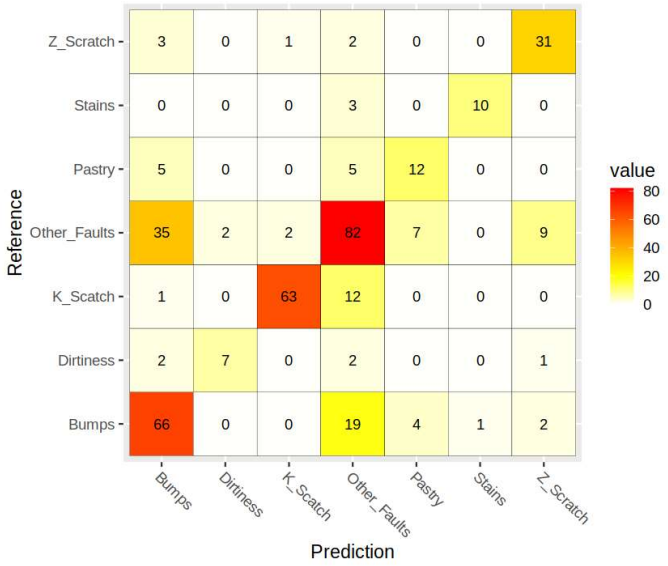
Without PCA



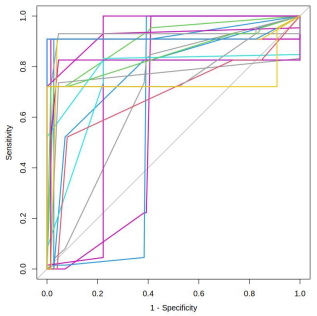
Accuracy: Train: 0.610, Test: 0.612



PCA

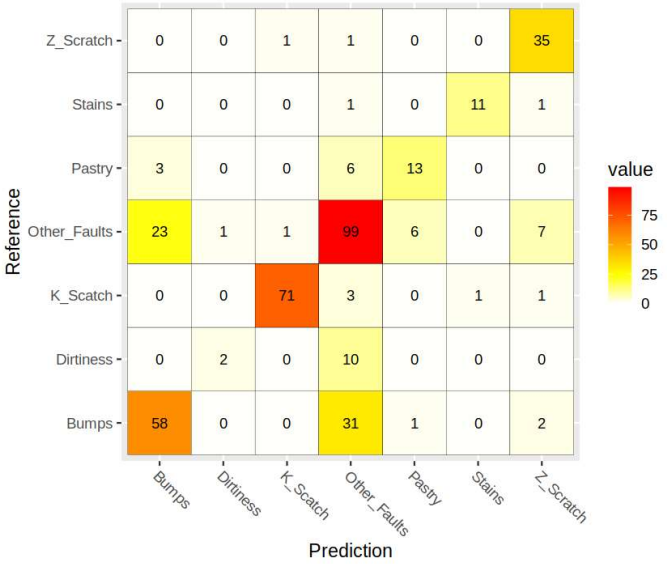


Accuracy: Train: 0.700, Test: 0.697

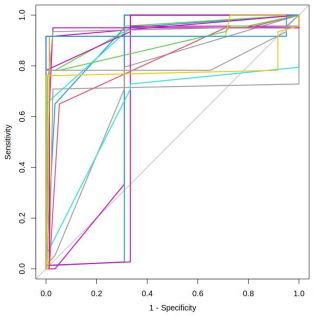


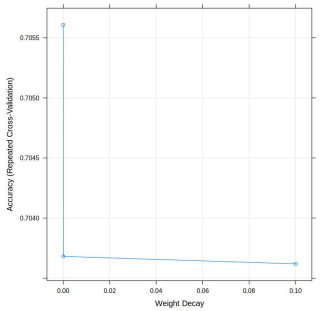
Multinomial Logistic Regression (10-fold CV)

Without PCA (decay = 0)

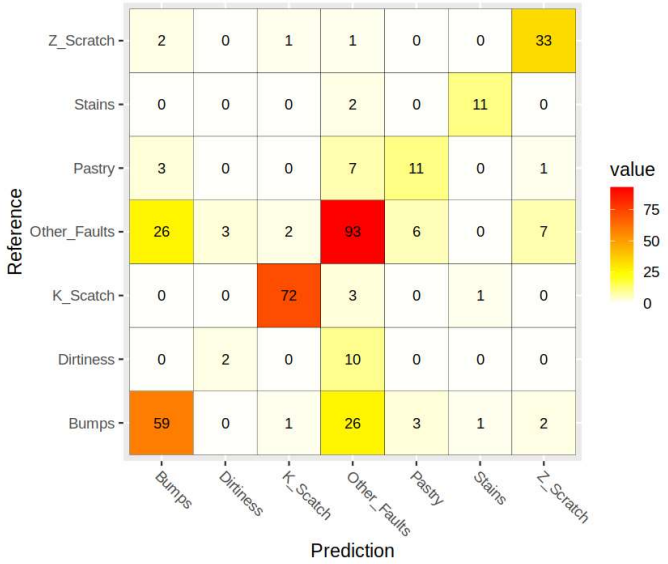


Accuracy: Train: 0.740, Test: 0.743

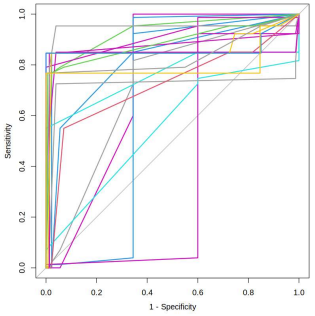


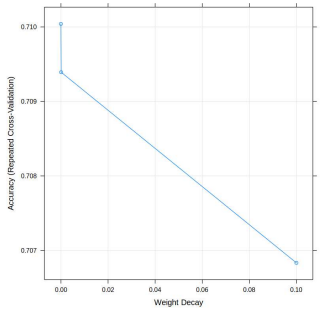


PCA (decay = 0)



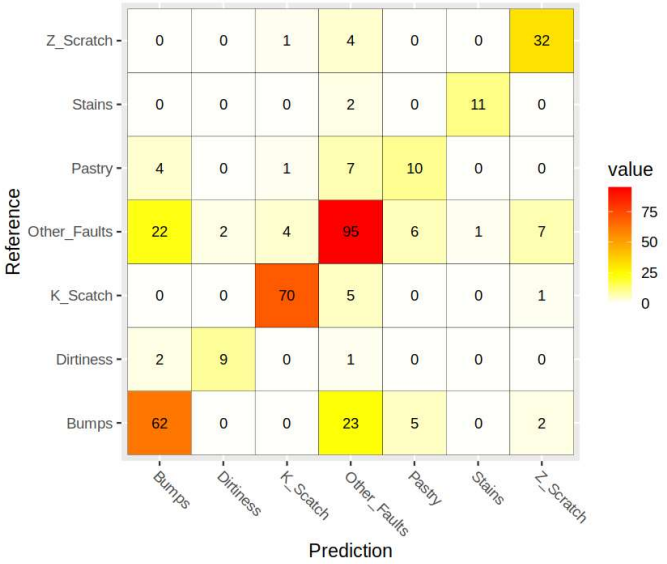
Accuracy: Train: 0.719, Test: 0.722



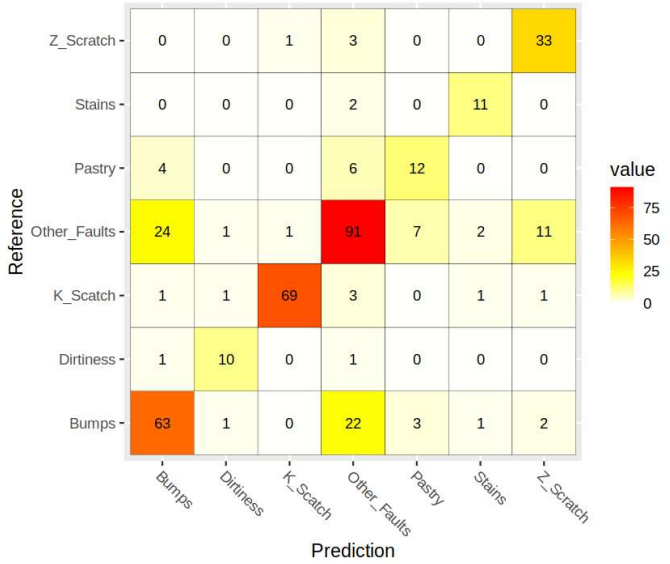


K-Nearest Neighbour (10-fold CV)

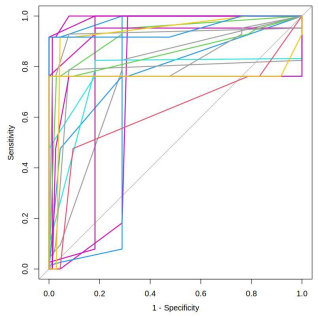
Without PCA (K = 1)

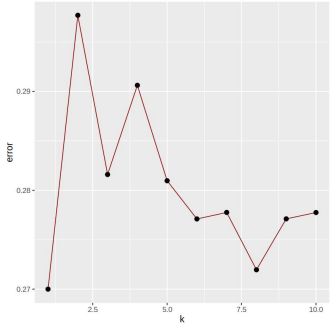


PCA (K = 5)

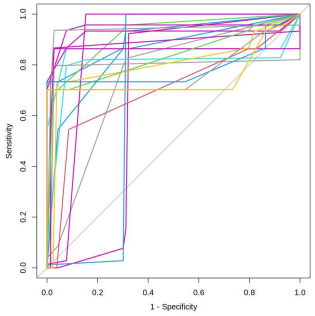


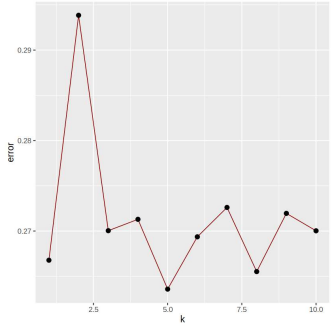
Accuracy: Train: 1, Test: 0.743





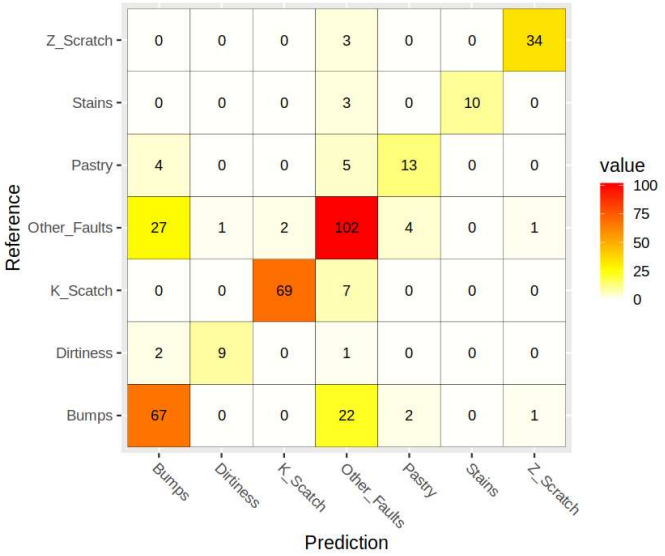
Accuracy: Train: 0.823, Test: 0.743



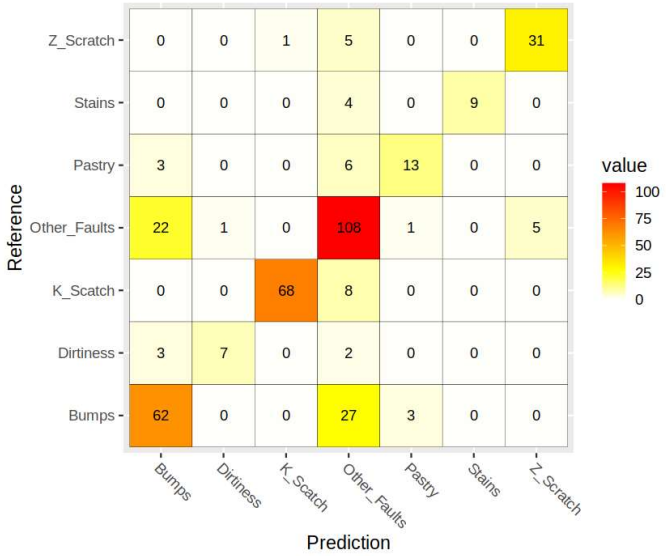


Random Forest (10-fold CV)

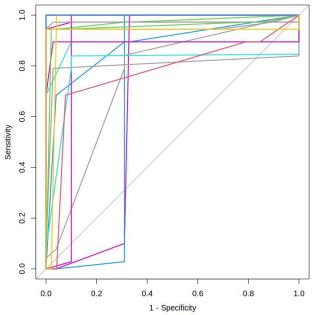
Without PCA (mtry = 27)

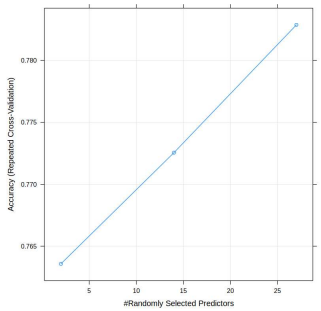


PCA (mtry = 7)

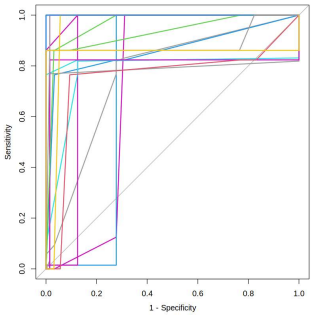


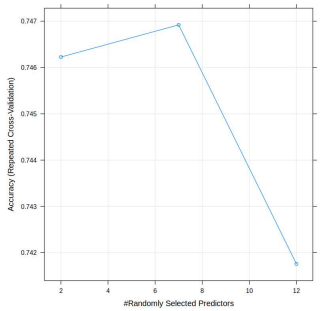
Accuracy: Train: 1, Test: 0.781





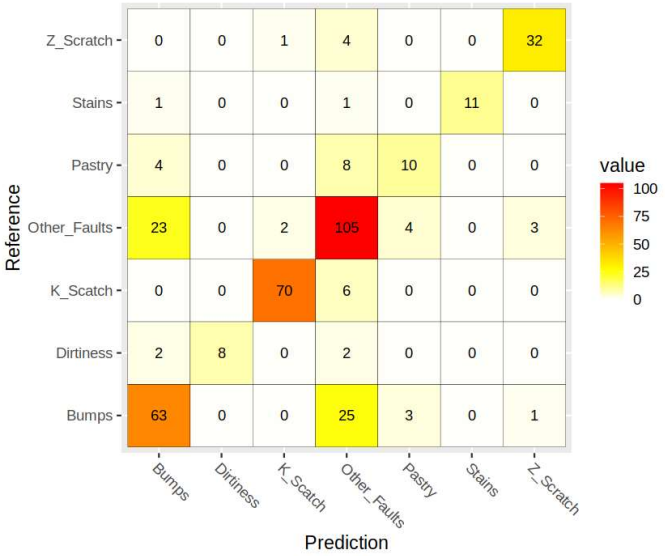
Accuracy: Train: 1, Test: 0.766



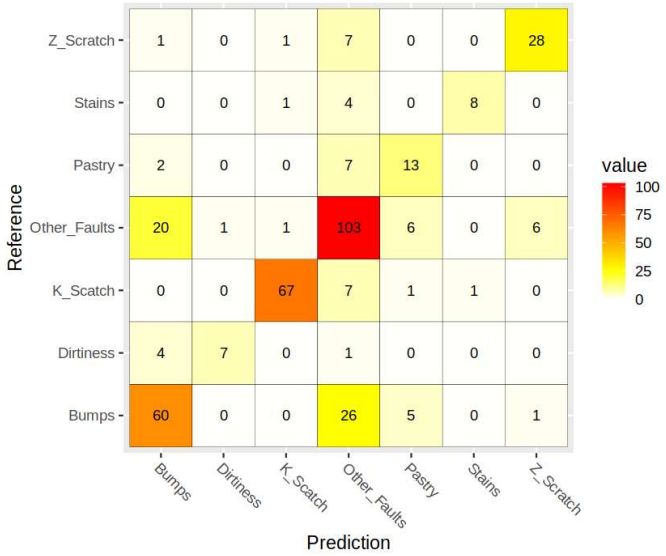


Boosting (10-fold CV)

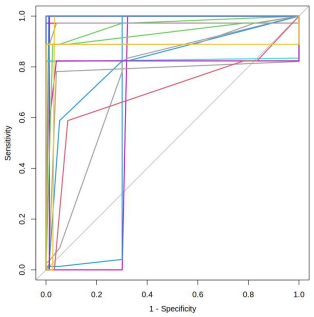
Without PCA (ntree = 150, depth = 3)

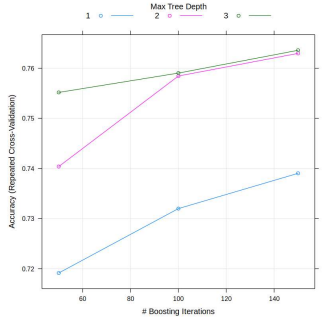


PCA (ntree = 150, depth = 3)

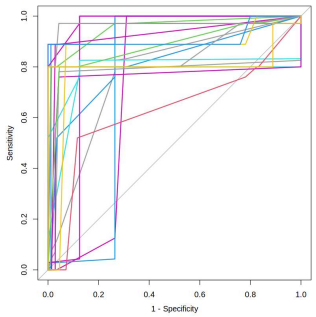


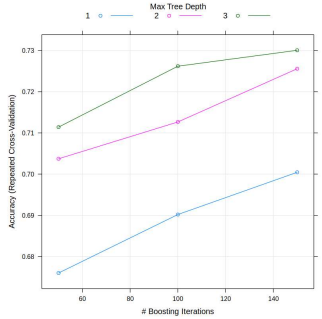
Accuracy: Train: 0.957, Test: 0.769





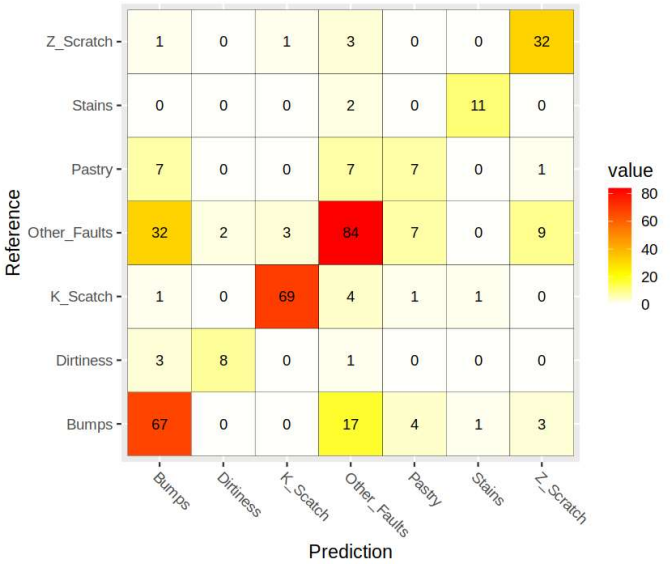
Accuracy: Train: 0.933, Test: 0.735



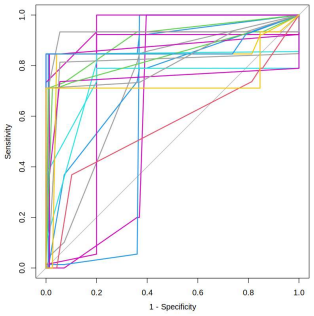


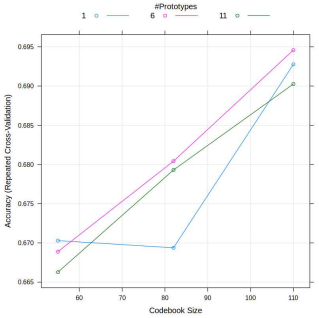
Learning Vector Quantization (10-fold CV)

Without PCA (size = 110, K = 6)

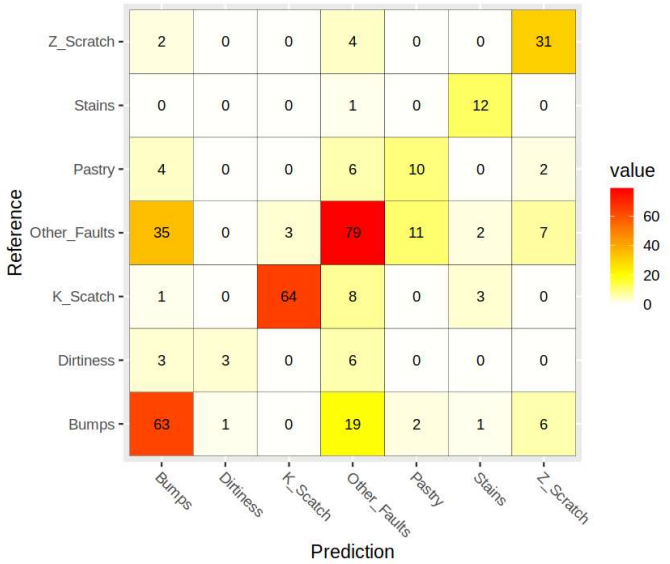


Accuracy: Train: 0.771, Test: 0.715

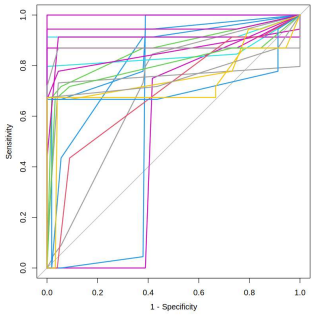


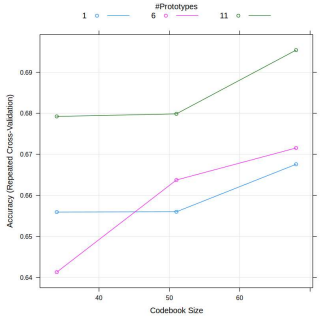


PCA (size = 68, K = 11)



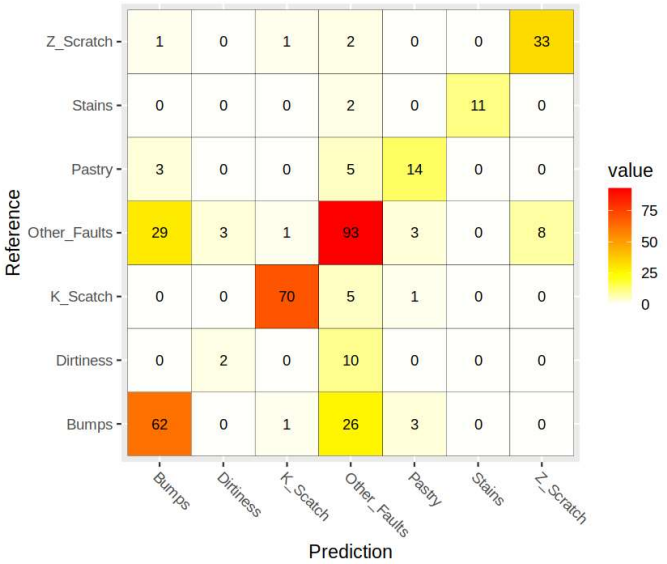
Accuracy: Train: 0.696, Test: 0.674



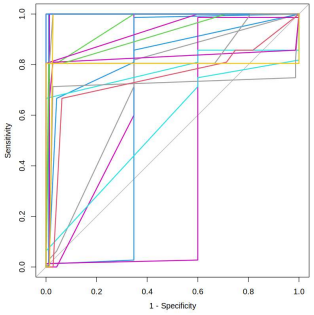


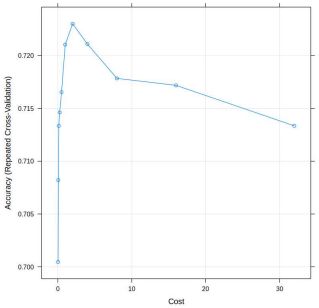
Support Vector Machine Linear Kernel (10-fold CV)

Without PCA (C = 2)

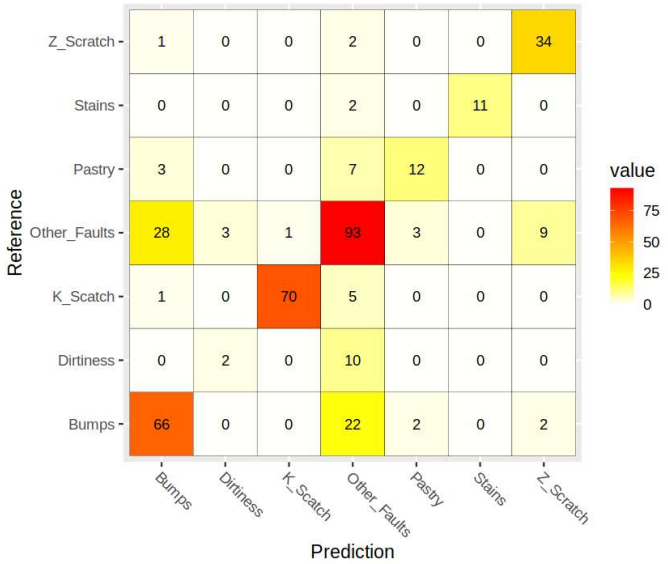


Accuracy: Train: 0.747, Test: 0.733

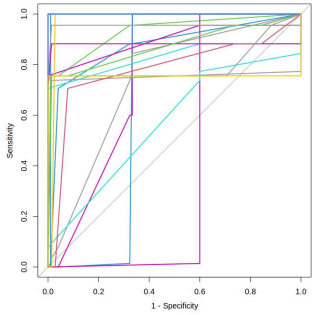


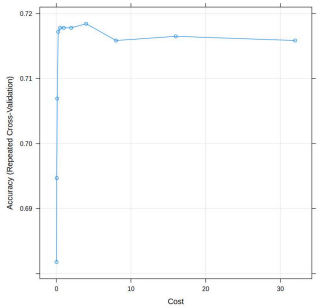


PCA (C = 4)



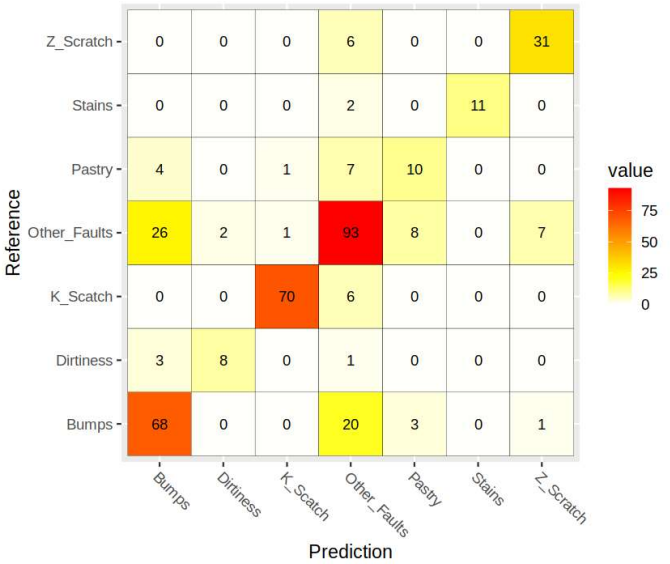
Accuracy: Train: 0.747, Test: 0.740



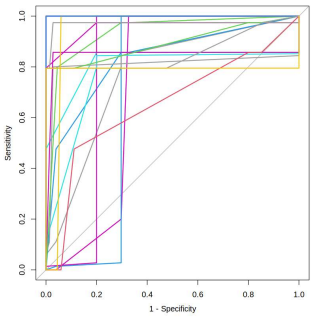


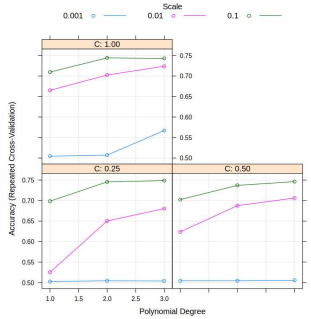
Support Vector Machine Polynomial Kernel (10-fold CV)

Without PCA (d = 3, scale = 0.1, C = 0.25)

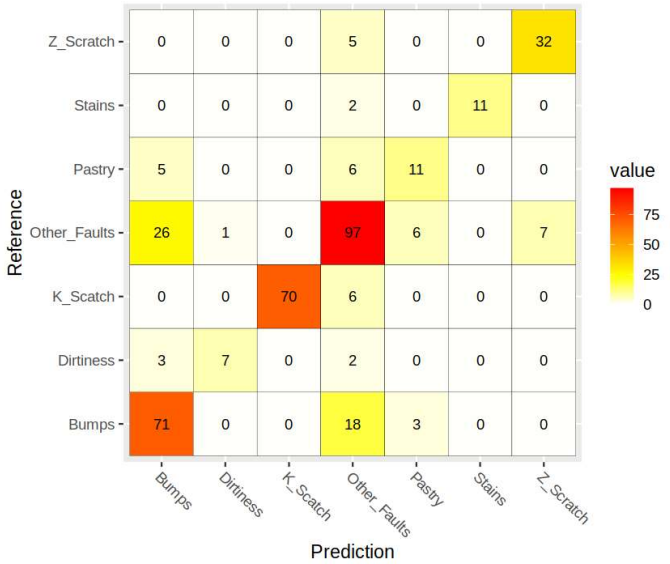


Accuracy: Train: 0.878, Test: 0.748

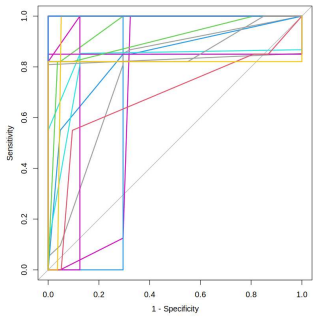


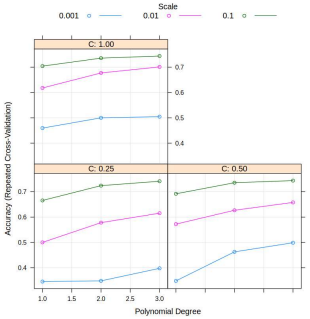


PCA (d = 3, scale = 0.1, C = 0.5)



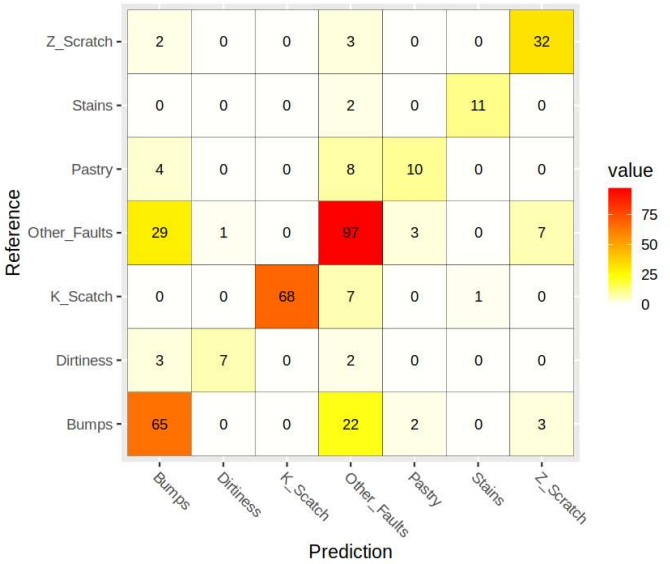
Accuracy: Train: 0.838, Test: 0.769



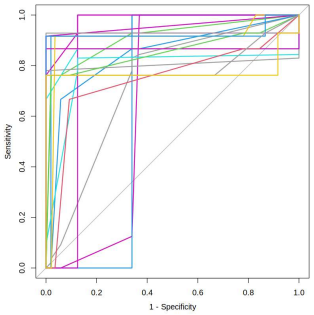


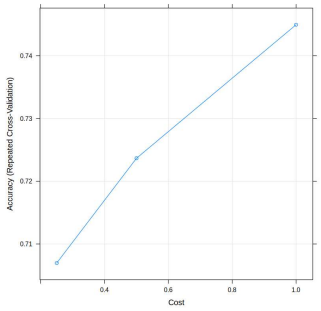
Support Vector Machine Radial Kernel (10-fold CV)

Without PCA (C = 1, sigma = 0.0338)

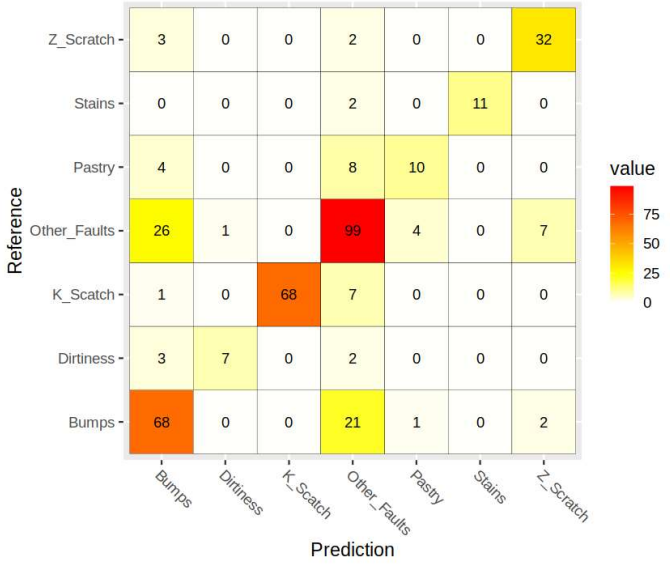


Accuracy: Train: 0.798, Test: 0.746

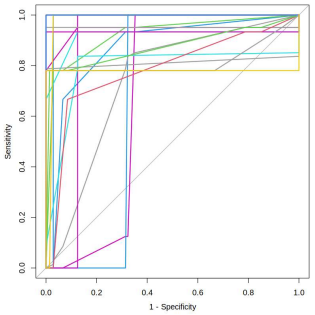


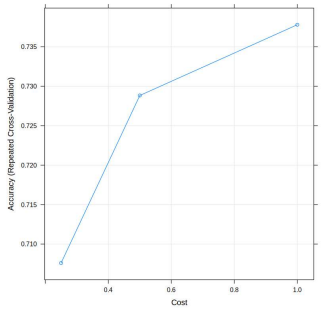


PCA (C = 1, sigma = 0.0647)



Accuracy: Train: 0.789, Test: 0.758





Conclusion and Discussion

Training Set Accuracy

Model	Without PCA	With PCA
GNB	0.6101804	0.7003866
MLG	0.7403351	0.7197165
KNN	1	0.8228093
RF	1	1
BOOST	0.9574742	0.9329897
LVQ	0.7712629	0.6958763
SVM Linear	0.7474227	0.7474227
SVM Polynomial	0.8782216	0.8376289
SVM Radial	0.7976804	0.7893041
Best	KNN and RF	RF
Best overall	RF	

Testing Set Accuracy

Model	Without PCA	With PCA
GNB	0.6118252	0.6966581
MLG	0.7429306	0.722365
KNN	0.7429306	0.7429306
RF	0.781491	0.7660668
BOOST	0.7686375	0.7352185
LVQ	0.714653	0.6735219
SVM Linear	0.7326478	0.7403599
SVM Polynomial	0.748072	0.7686375
SVM Radial	0.7455013	0.7583548
Best	RF	SVM Polynomial
Best overall	RF without PCA	