# Methodology:

1. 38 features extracted as follows:
   1. GLRLMS {mean of 4 angles} (11)
   2. GLCM (20)
   3. Hu moments (7)
2. “Low” and “High” grades lumped
3. Various feature selection methods applied:
   1. Univariate selection – top k-percentile features selected
4. 10-fold cross validation on final feature list

## Univariate selection

Top k-percentile features selected using various scoring functions:

1. Chi-squared distance
2. ANOVA F-value
3. Mutual information

### Chi-squared distance

Percentiles selected: top 10%-70%

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Trial No. | Percentile | # variables selected | Accuracy | Best classifier |
| 1 | Top 10 percentile | 4/38 | 79.2% | Weighted KNN |
| 2 | 79.2% | Weighted KNN |
| 3 | 76.6% | Weighted KNN & Quadratic SVM |
| 1 | Top 20 percentile | 8/38 | 81.8% | Boosted Trees |
| 2 | 76.6% | Gaussian SVM |
| 3 | 77.9% | Gaussian SVM |
| 1 | Top 30 percentile | 12/38 | 85.7% | Fine KNN |
| 2 | 88.3% | Fine KNN |
| 3 | 88.3% | Fine KNN |
| 1 | Top 40 percentile | 15/38 | 85.7% | Fine KNN |
| 2 | 87% | Fine KNN |
| 3 | 88.3% | Weighted KNN |
| 1 | Top 50 percentile | 19/38 | 87% | Fine KNN |
| 2 | 88.3% | Fine KNN |
| 3 | 87% | Fine KNN |
| 1 | Top 60 percentile | 23/38 | 87% | Fine KNN |
| 2 | 85.7% | Fine KNN |
| 3 | 87% | Fine KNN |
| 1 | Top 70 percentile | 26/38 | 94.8% | Fine KNN |
| 2 | 93.5% | Fine KNN |
| 3 | 90.9% | Fine KNN |

### ANOVA F-value

Percentiles selected: top 10%-70%

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Trial No. | Percentile | # variables selected | Accuracy | Best classifier |
| 1 | Top 10 percentile | 4/38 | 76.6% | Weighted KNN |
| 2 | 80.5% | Decision Trees |
| 3 | 77.9% | Decision Trees |
| 1 | Top 20 percentile | 8/38 | 79.2% | Fine KNN |
| 2 | 81.8% | Weighted KNN |
| 3 | 81.8% | Weighted KNN |
| 1 | Top 30 percentile | 12/38 | 80.5% | Decision Trees |
| 2 | 79.2% | Decision Trees |
| 3 | 81.8% | Fine KNN |
| 1 | Top 40 percentile | 15/38 | 85.7% | Fine KNN & Gaussian SVM |
| 2 | 87% | Fine KNN |
| 3 | 85.7% | Fine KNN & Gaussian SVM |
| 1 | Top 50 percentile | 19/38 | 83.1% | Fine KNN |
| 2 | 84.4% | Fine KNN |
| 3 | 83.1% | Fine KNN |
| 1 | Top 60 percentile | 23/38 | 80.5% | Weighted KNN & Gaussian SVM |
| 2 | 84.4% | Gaussian SVM |
| 3 | 83.1% | Bagged Trees |
| 1 | Top 70 percentile | 26/38 | 81.8% | Gaussian SVM |
| 2 | 84.4% | Bagged trees |
| 3 | 84.4% | Gaussian SVM |

### Mutual Information

Percentiles selected: top 10%-70%

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Trial No. | Percentile | # variables selected | Accuracy | Best classifier |
| 1 | Top 10 percentile | 4/38 | 79.2% | Decision Trees |
| 2 | 76.6% | Decision Trees |
| 3 | 80.5% | Bagged Trees |
| 1 | Top 20 percentile | 8/38 | 84.4% | Weighted KNN |
| 2 | 79.9% | Weighted KNN |
| 3 | 80.5% | Weighted KNN |
| 1 | Top 30 percentile | 12/38 | 90.9% | Weighted KNN |
| 2 | 88.3% | Weighted KNN |
| 3 | 90.9% | Weighted KNN |
| 1 | Top 40 percentile | 15/38 | 84.4% | Weighted KNN |
| 2 | 84.4% | Weighted & Fine KNN |
| 3 | 87% | Weighted KNN |
| 1 | Top 50 percentile | 19/38 | 84.4% | Fine KNN |
| 2 | 84.4% | Weighted KNN |
| 3 | 81.8% | Fine & Weighted KNN |
| 1 | Top 60 percentile | 23/38 | 84.4% | Fine KNN |
| 2 | 83.1% | Weighted KNN |
| 3 | 84.4% | Fine & Weighted KNN |
| 1 | Top 70 percentile | 26/38 | 84.4% | Fine % Weighted KNN |
| 2 | 85.7% | Fine KNN |
| 3 | 88.3% | Fine KNN |