

Supplementary Material

Additional results are provided below for each of the application domains examined in the main paper:

- cancer classification from genomic data (Table S1 and Table S2)
- text categorization (Table S3 and Table S4)
- image classification (Table S5 and Table S6)

Table S1. *Uterus* and *Omentum* datasets: *F-measure* and *G-mean* performance of the *RF* classifier in conjunction with data balancing (*RUS/SMOTE*) and feature selection (1% of the original features, as selected by the *IG* ranker).

| <i>Dataset</i> | <i>Performance measure</i> | <i>Baseline</i> | <i>FS alone</i> | <i>FS with RUS (3:1)</i> | <i>FS with RUS (2:1)</i> | <i>FS with RUS (1:1)</i> | <i>FS with SMOTE (3:1)</i> | <i>FS with SMOTE (2:1)</i> | <i>FS with SMOTE (1:1)</i> | |
|----------------|----------------------------|-----------------|-----------------|--------------------------|--------------------------|--------------------------|----------------------------|----------------------------|----------------------------|---------------------------|
| <i>Uterus</i> | <i>F-measure</i> | 0.26 | 0.50 | 0.62 | 0.61 | 0.55 | 0.62 | 0.64 | 0.65 | <i>FS before sampling</i> |
| | | | | 0.62 | 0.60 | 0.53 | 0.62 | 0.63 | 0.63 | <i>Sampling before FS</i> |
| | <i>G-mean</i> | 0.38 | 0.63 | 0.84 | 0.87 | 0.88 | 0.80 | 0.83 | 0.87 | <i>FS before sampling</i> |
| | | | | 0.84 | 0.86 | 0.87 | 0.80 | 0.82 | 0.85 | <i>Sampling before FS</i> |
| <i>Omentum</i> | <i>F-measure</i> | 0.00 | 0.32 | 0.54 | 0.51 | 0.46 | 0.53 | 0.54 | 0.56 | <i>FS before sampling</i> |
| | | | | 0.53 | 0.50 | 0.44 | 0.55 | 0.56 | 0.55 | <i>Sampling before FS</i> |
| | <i>G-mean</i> | 0.00 | 0.47 | 0.84 | 0.86 | 0.87 | 0.74 | 0.75 | 0.79 | <i>FS before sampling</i> |
| | | | | 0.83 | 0.86 | 0.87 | 0.76 | 0.79 | 0.81 | <i>Sampling before FS</i> |

Table S2. *Uterus* and *Omentum* datasets: *F-measure* and *G-mean* performance of the *RF* classifier in conjunction with cost-sensitive methods (*MinCost/Weighting*) and feature selection (1% of the original features, as selected by the *IG* ranker).

| <i>Dataset</i> | <i>Performance measure</i> | <i>Baseline</i> | <i>FS alone</i> | <i>FS with MinCost (2)</i> | <i>FS with MinCost (3)</i> | <i>FS with MinCost (4)</i> | <i>FS with Weighting (2)</i> | <i>FS with Weighting (3)</i> | <i>FS with Weighting (4)</i> |
|----------------|----------------------------|-----------------|-----------------|----------------------------|----------------------------|----------------------------|------------------------------|------------------------------|------------------------------|
| <i>Uterus</i> | <i>F-measure</i> | 0.26 | 0.50 | 0.62 | 0.64 | 0.63 | 0.58 | 0.60 | 0.60 |
| | <i>G-mean</i> | 0.38 | 0.63 | 0.80 | 0.86 | 0.88 | 0.72 | 0.75 | 0.75 |
| <i>Omentum</i> | <i>F-measure</i> | 0.00 | 0.32 | 0.52 | 0.55 | 0.57 | 0.40 | 0.45 | 0.44 |
| | <i>G-mean</i> | 0.00 | 0.47 | 0.70 | 0.78 | 0.83 | 0.56 | 0.61 | 0.61 |

Table S3. *Trade* and *Interest* datasets: *F-measure* and *G-mean* performance of the *RF* classifier in conjunction with data balancing (*RUS/SMOTE*) and feature selection (5% of the original features, as selected by the *IG* ranker).

| <i>Dataset</i> | <i>Performance measure</i> | <i>Baseline</i> | <i>FS alone</i> | <i>FS with RUS (3:1)</i> | <i>FS with RUS (2:1)</i> | <i>FS with RUS (1:1)</i> | <i>FS with SMOTE (3:1)</i> | <i>FS with SMOTE (2:1)</i> | <i>FS with SMOTE (1:1)</i> | |
|-----------------|----------------------------|-----------------|-----------------|--------------------------|--------------------------|--------------------------|----------------------------|----------------------------|----------------------------|---------------------------|
| <i>Trade</i> | <i>F-measure</i> | 0.31 | 0.63 | 0.69 | 0.62 | 0.48 | 0.68 | 0.67 | 0.66 | <i>FS before sampling</i> |
| | | | | 0.68 | 0.60 | 0.46 | 0.61 | 0.63 | 0.63 | <i>Sampling before FS</i> |
| | <i>G-mean</i> | 0.43 | 0.69 | 0.93 | 0.94 | 0.94 | 0.74 | 0.74 | 0.73 | <i>FS before sampling</i> |
| | | | | 0.93 | 0.95 | 0.94 | 0.68 | 0.69 | 0.69 | <i>Sampling before FS</i> |
| <i>Interest</i> | <i>F-measure</i> | 0.53 | 0.62 | 0.69 | 0.64 | 0.53 | 0.67 | 0.67 | 0.68 | <i>FS before sampling</i> |
| | | | | 0.69 | 0.63 | 0.51 | 0.66 | 0.66 | 0.65 | <i>Sampling before FS</i> |
| | <i>G-mean</i> | 0.62 | 0.71 | 0.93 | 0.94 | 0.95 | 0.76 | 0.76 | 0.77 | <i>FS before sampling</i> |
| | | | | 0.93 | 0.95 | 0.95 | 0.74 | 0.75 | 0.74 | <i>Sampling before FS</i> |

Table S4. *Trade* and *Interest* datasets: *F-measure* and *G-mean* performance of the *RF* classifier in conjunction with cost-sensitive methods (*MinCost/Weighting*) and feature selection (5% of the original features, as selected by the *IG* ranker).

| <i>Dataset</i> | <i>Performance measure</i> | <i>Baseline</i> | <i>FS alone</i> | <i>FS with MinCost (2)</i> | <i>FS with MinCost (3)</i> | <i>FS with MinCost (4)</i> | <i>FS with Weighting (2)</i> | <i>FS with Weighting (3)</i> | <i>FS with Weighting (4)</i> |
|-----------------|----------------------------|-----------------|-----------------|----------------------------|----------------------------|----------------------------|------------------------------|------------------------------|------------------------------|
| <i>Trade</i> | <i>F-measure</i> | 0.31 | 0.63 | 0.75 | 0.75 | 0.72 | 0.70 | 0.69 | 0.71 |
| | <i>G-mean</i> | 0.43 | 0.69 | 0.84 | 0.90 | 0.92 | 0.76 | 0.76 | 0.77 |
| <i>Interest</i> | <i>F-measure</i> | 0.53 | 0.62 | 0.72 | 0.74 | 0.73 | 0.67 | 0.68 | 0.70 |
| | <i>G-mean</i> | 0.62 | 0.71 | 0.82 | 0.89 | 0.92 | 0.76 | 0.78 | 0.79 |

Table S5. Mountain and Urban datasets: *F-measure* and *G-mean* performance of the RF classifier in conjunction with data balancing (RUS/SMOTE) and feature selection (CFS filter).

| <i>Dataset</i> | <i>Performance measure</i> | <i>Baseline</i> | <i>FS alone</i> | <i>FS with RUS (3:1)</i> | <i>FS with RUS (2:1)</i> | <i>FS with RUS (1:1)</i> | <i>FS with SMOTE (3:1)</i> | <i>FS with SMOTE (2:1)</i> | <i>FS with SMOTE (1:1)</i> | |
|-----------------|----------------------------|-----------------|-----------------|--------------------------|--------------------------|--------------------------|----------------------------|----------------------------|----------------------------|---------------------------|
| <i>Mountain</i> | <i>F-measure</i> | 0.51 | 0.55 | 0.59 | 0.63 | 0.61 | 0.57 | 0.62 | 0.64 | <i>FS before sampling</i> |
| | | | | 0.58 | 0.63 | 0.60 | 0.57 | 0.61 | 0.64 | <i>Sampling before FS</i> |
| | <i>G-mean</i> | 0.60 | 0.64 | 0.68 | 0.75 | 0.78 | 0.66 | 0.72 | 0.78 | <i>FS before sampling</i> |
| | | | | 0.67 | 0.75 | 0.77 | 0.66 | 0.72 | 0.77 | <i>Sampling before FS</i> |
| <i>Urban</i> | <i>F-measure</i> | 0.56 | 0.58 | 0.64 | 0.66 | 0.61 | 0.64 | 0.66 | 0.67 | <i>FS before sampling</i> |
| | | | | 0.64 | 0.66 | 0.61 | 0.64 | 0.66 | 0.67 | <i>Sampling before FS</i> |
| | <i>G-mean</i> | 0.64 | 0.67 | 0.75 | 0.80 | 0.82 | 0.74 | 0.77 | 0.81 | <i>FS before sampling</i> |
| | | | | 0.75 | 0.81 | 0.83 | 0.73 | 0.77 | 0.81 | <i>Sampling before FS</i> |

Table S6. Mountain and Urban datasets: *F-measure* and *G-mean* performance of the RF classifier in conjunction with cost-sensitive methods (*MinCost/Weighting*) and feature selection (CFS filter).

| <i>Dataset</i> | <i>Performance measure</i> | <i>Baseline</i> | <i>FS alone</i> | <i>FS with MinCost (2)</i> | <i>FS with MinCost (3)</i> | <i>FS with MinCost (4)</i> | <i>FS with Weighting (2)</i> | <i>FS with Weighting (3)</i> | <i>FS with Weighting (4)</i> |
|-----------------|----------------------------|-----------------|-----------------|----------------------------|----------------------------|----------------------------|------------------------------|------------------------------|------------------------------|
| <i>Mountain</i> | <i>F-measure</i> | 0.51 | 0.55 | 0.64 | 0.61 | 0.57 | 0.62 | 0.63 | 0.64 |
| | <i>G-mean</i> | 0.60 | 0.64 | 0.78 | 0.79 | 0.76 | 0.72 | 0.75 | 0.77 |
| <i>Urban</i> | <i>F-measure</i> | 0.56 | 0.58 | 0.66 | 0.64 | 0.61 | 0.65 | 0.66 | 0.66 |
| | <i>G-mean</i> | 0.64 | 0.67 | 0.80 | 0.83 | 0.83 | 0.75 | 0.77 | 0.79 |