Supplementary Table 1. Calculation of Area under ROC, Sensitivity and Specificity of Six Statistical Models

Models: CART, SVM, plr, nnet, knn, rrf

Number of resamples: 50

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ROC |  |  |  |  |  |  |  |
|  | Min. | 1stq | Median | Mean | 3rd | Max. | NA's |
| CART | 0.75 | 0.8571 | 0.9167 | 0.9098 | 0.9286 | 1 | 0 |
| SVM | 0.881 | 1 | 1 | 0.985 | 1 | 1 | 0 |
| plr | 0.6905 | 0.8571 | 0.9388 | 0.9218 | 1 | 1 | 0 |
| nnet | 0.7857 | 0.898 | 0.9286 | 0.9323 | 1 | 1 | 0 |
| knn | 0.7976 | 0.8776 | 0.9691 | 0.9373 | 1 | 1 | 0 |
| rrf | 0.8333 | 0.9388 | 1 | 0.9638 | 1 | 1 | 0 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Specificity |  |  |  |  |  |  |  |
|  | Min. | 1stq | Median | Mean | 3rd | Max. | NA's |
| CART | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| SVM | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| plr | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| nnet | 0.7143 | 1 | 1 | 0.9657 | 1 | 1 | 0 |
| knn | 0.8571 | 1 | 1 | 0.9914 | 1 | 1 | 0 |
| rrf | 1 | 1 | 1 | 1 | 1 | 1 | 0 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Sensitivity |  |  |  |  |  |  |  |
|  | Min. | 1stq | Median | Mean | 3rd | Max. | NA's |
| CART | 0.75 | 0.8571 | 0.9167 | 0.9098 | 0.9286 | 1 | 0 |
| SVM | 0.881 | 1 | 1 | 0.985 | 1 | 1 | 0 |
| plr | 0.6905 | 0.8571 | 0.9388 | 0.9218 | 1 | 1 | 0 |
| nnet | 0.7857 | 0.898 | 0.9286 | 0.9323 | 1 | 1 | 0 |
| knn | 0.7976 | 0.8776 | 0.9691 | 0.9373 | 1 | 1 | 0 |
| rrf | 0.8333 | 0.9388 | 1 | 0.9638 | 1 | 1 | 0 |

Supplementary Table 2. The Oral Cancer Risk Index (OCRI) of OLK Patients

刘瑶，这里加3个column，就是原来脱落细胞报告的结果（positive, atypical, negative）,随访时间，随访结果。OCRI改成小数点后两位数

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Case | Age | Sex | Site a | Smoking | Drinking | Clinical  classification | Pathological  diagnosis | Exfoliative cytology | | Clinical follow-up | |
| Qualitative b | OCRI | Duration (mo) | Outcome c |
| 128111 | 50 | F | T | N | N | Heterogeneous | mild dysplasia |  | 0.593673 |  |  |
| 128113 | 55 | F | G | Y | N | Heterogeneous | mild dysplasia |  | 0.005427 |  |  |
| 128114 | 57 | F | T | N | N | Homogeneous | moderate dysplasia |  | 0.260392 |  |  |
| 128116 | 52 | M | T | Y | Y | Homogeneous | mild dysplasia |  | 0.079024 |  |  |
| 128119 | 57 | M | G | Y | N | Heterogeneous | no dysplasia |  | 0.067476 |  |  |
| 128125 | 50 | F | T | N | N | Homogeneous | moderate dysplasia |  | 0.006794 |  |  |
| 128126 | 64 | F | G | Y | N | Homogeneous | no dysplasia |  | 0.013385 |  |  |
| 128128 | 47 | M | G | N | Y | Homogeneous | moderate dysplasia |  | 0.054023 |  |  |
| 128129 | 62 | F | G | N | N | Homogeneous | no dysplasia |  | 0.118226 |  |  |
| 128131 | 68 | F | B | N | N | Homogeneous | mild dysplasia |  | 0.640342 |  |  |
| 128132 | 52 | M | B | N | N | Homogeneous | mild dysplasia |  | 0.390436 |  |  |
| 128133 | 46 | F | B | N | N | Homogeneous | no dysplasia |  | 0.006622 |  |  |
| 128134 | 68 | F | G | N | N | Homogeneous | moderate dysplasia |  | 0.13195 |  |  |
| 128136 | 46 | F | G | N | N | Homogeneous | no dysplasia |  | 0.061511 |  |  |
| 128138 | 55 | M | T | Y | N | Homogeneous | moderate dysplasia |  | 0.068795 |  |  |
| 128140 | 60 | F | T | N | N | Heterogeneous | moderate dysplasia |  | 0.108728 |  |  |
| 128141 | 48 | F | T | N | N | Homogeneous | mild dysplasia |  | 0.884359 | 40 | OSCC |
| 128145 | 65 | M | G | N | Y | Homogeneous | no dysplasia |  | 0.158812 |  |  |
| 128147 | 44 | F | B | N | N | Homogeneous | no dysplasia |  | 0.498427 |  |  |
| 128149 | 52 | F | G | N | N | Homogeneous | moderate dysplasia |  | 0.002838 |  |  |
| 128150 | 62 | F | T | N | N | Heterogeneous | no dysplasia |  | 0.102624 |  |  |
| 9449998 | 71 | F | T | N | N | Homogeneous | moderate dysplasia |  | 0.038013 |  |  |
| 9449999 | 38 | F | G | N | N | Homogeneous | moderate dysplasia |  | 0.070791 |  |  |
| 9450000 | 71 | F | G | N | N | Homogeneous | no dysplasia |  | 0.232334 |  |  |
| 9558745 | 67 | M | B | N | N | Heterogeneous | moderate dysplasia |  | 0.065335 |  |  |
| 9559194 | 75 | M | B | Y | N | Homogeneous | mild dysplasia |  | 0.1005 |  |  |
| 9559197 | 40 | F | B | N | N | Homogeneous | no dysplasia |  | 0.099049 |  |  |
| 9559198 | 40 | F | G | N | N | Homogeneous | mild dysplasia |  | 0.032038 |  |  |
| 9559200 | 40 | F | T | N | N | Homogeneous | mild dysplasia |  | 0.072301 |  |  |
| 9559201 | 70 | M | T | Y | Y | Homogeneous | mild dysplasia |  | 0.002368 |  |  |
| 9559202 | 70 | M | T | Y | Y | Heterogeneous | moderate dysplasia |  | 0.002117 |  |  |
| 9559231 | 48 | M | B | N | Y | Homogeneous | moderate dysplasia |  | 0.082328 |  |  |
| 9559232 | 76 | F | T | N | N | Homogeneous | moderate dysplasia |  | 0.080773 |  |  |
| 9559233 | 59 | M | B | Y | Y | Homogeneous | mild dysplasia |  | 0.03589 |  |  |
| 9559234 | 70 | F | G | N | N | Homogeneous | moderate dysplasia |  | 0.097379 |  |  |
| 9559235 | 70 | M | G | Y | N | Homogeneous | moderate dysplasia |  | 0.101886 |  |  |
| 9559238 | 58 | M | G | Y | Y | Homogeneous | mild dysplasia |  | 0.02133 |  |  |
| 9559239 | 60 | F | T | N | N | Homogeneous | moderate dysplasia |  | 0.019513 |  |  |
| 9559240 | 68 | M | B | Y | N | Homogeneous | moderate dysplasia |  | 0.086112 |  |  |
| 9559241 | 72 | F | T | N | N | Homogeneous | mild dysplasia |  | 0.929742 |  |  |
| 9559246 | 67 | F | B | N | N | Heterogeneous | no dysplasia |  | 0.906709 |  |  |
| 9559247 | 67 | F | B | N | N | Homogeneous | no dysplasia |  | 0.969823 |  |  |
| 9559248 | 59 | F | G | N | N | Homogeneous | mild dysplasia |  | 0.08105 |  |  |
| 9559250 | 54 | F | G | N | N | Homogeneous | mild dysplasia |  | 0.02287 |  |  |
| 9559261 | 52 | M | P | Y | Y | Heterogeneous | moderate dysplasia |  | 0.467263 |  |  |
| 9559265 | 67 | M | B | Y | Y | Homogeneous | moderate dysplasia |  | 0.004175 |  |  |
| 9559267 | 57 | M | G | N | N | Homogeneous | moderate dysplasia |  | 0.918019 |  |  |
| 9559272 | 67 | M | B | Y | Y | Heterogeneous | no dysplasia |  | 0.384193 |  |  |
| 9559275 | 52 | F | G | N | N | Homogeneous | mild dysplasia |  | 0.007302 |  |  |
| 9559277 | 54 | F | G | N | N | Homogeneous | no dysplasia |  | 0.011013 |  |  |
| 9559283 | 85 | M | P | N | Y | Homogeneous | no dysplasia |  | 0.863304 |  |  |
| 9559286 | 79 | F | G | Y | N | Homogeneous | moderate dysplasia |  | 0.071386 |  |  |
| 9559287 | 61 | F | G | Y | N | Homogeneous | mild dysplasia |  | 0.016794 |  |  |
| 9559288 | 54 | F | G | N | N | Homogeneous | moderate dysplasia |  | 0.023508 |  |  |
| 9559289 | 71 | F | G | N | N | Homogeneous | mild dysplasia |  | 0.025857 |  |  |
| 9559290 | 25 | M | T | Y | N | Heterogeneous | moderate dysplasia |  | 0.070226 |  |  |
| 9559291 | 58 | F | B | N | N | Homogeneous | no dysplasia |  | 0.028518 |  |  |
| 9559292 | 65 | F | P | N | N | Homogeneous | no dysplasia |  | 0.041541 |  |  |
| 9559295 | 52 | M | B | Y | N | Homogeneous | mild dysplasia |  | 0.012601 |  |  |
| 9559296 | 68 | M | B | N | Y | Homogeneous | no dysplasia |  | 0.024196 |  |  |
| 9559297 | 60 | F | T | N | N | Heterogeneous | moderate dysplasia |  | 0.065405 |  |  |
| 9559298 | 43 | M | T | Y | Y | Homogeneous | mild dysplasia |  | 0.016055 |  |  |
| 9559398 | 60 | M | B | Y | N | Homogeneous | mild dysplasia |  | 0.19465 |  |  |
| 9559412 | 53 | F | T | N | N | Heterogeneous | mild dysplasia |  | 0.520302 |  |  |
| 9559430 | 48 | F | T | N | N | Heterogeneous | moderate dysplasia |  | 0.43501 |  |  |
| 9559482 | 67 | F | B | N | N | Homogeneous | mild dysplasia |  | 0.148199 |  |  |
| 9559486 | 55 | M | B | Y | N | Homogeneous | mild dysplasia |  | 0.060511 |  |  |
| 9559488 | 39 | M | G | N | N | Homogeneous | mild dysplasia |  | 0.145039 |  |  |
| 9559489 | 53 | F | T | N | N | Homogeneous | moderate dysplasia |  | 0.62345 |  |  |
| 9559490 | 60 | M | G | Y | Y | Heterogeneous | no dysplasia |  | 0.027713 |  |  |
| 9559491 | 47 | M | G | N | Y | Homogeneous | mild dysplasia |  | 0.608597 |  |  |
| 9559504 | 50 | M | B | Y | N | Homogeneous | mild dysplasia |  | 0.073852 |  |  |
| 9559505 | 74 | M | G | N | N | Heterogeneous | mild dysplasia |  | 0.066078 |  |  |
| 9559507 | 80 | F | B | N | N | Homogeneous | mild dysplasia |  | 0.135224 |  |  |
| 9559542 | 29 | M | B | Y | N | Homogeneous | mild dysplasia |  | 0.104602 |  |  |
| 9559546 | 65 | M | T | Y | Y | Homogeneous | mild dysplasia |  | 0.003294 |  |  |
| 9559547 | 63 | M | G | Y | Y | Heterogeneous | no dysplasia |  | 0.891315 |  |  |
| 9559550 | 66 | F | G | N | N | Heterogeneous | moderate dysplasia |  | 0.067471 |  |  |
| 9559561 | 57 | M | B | Y | N | Homogeneous | moderate dysplasia |  | 0.633512 |  |  |
| 9559565 | 44 | M | G | N | N | Homogeneous | no dysplasia |  | 0.028842 |  |  |
| 9559570 | 61 | F | G | N | N | Homogeneous | moderate dysplasia |  | 0.734662 |  |  |
| 9559237-1 | 58 | M | G | Y | Y | Homogeneous | mild dysplasia |  | 0.059771 |  |  |

a B: Buccal, T: Tongue, G: Gingival, P: Palate

b Negative, no aneupolid cells; Positive, definitive cellular evidence of epithelial dysplasia or carcinoma; Atypical, abnormal epithelial changes of uncertain diagnostic significance

c OSCC, OSCC was diagnosed; No, No cancer was diagnosed; Lost; Lost in follow-up; Death: Death due to other diseases