**Appendix 6: Features of datasets used for development and validation of AI models**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| AuthorID | Data source | Data type | Dataset size | Type of validation | % of training set | % of validation set | % of test set |
| Abbas36 | Public database | Radiology images | 196 | Training-test split | 70% | N/A | 30% |
| Abdelmageed64 | Public database | Biological data | N/A | N/A | N/A | N/A | N/A |
| Al-Qaness72 | Public database | Epidemiological Data | N/A | Training-test split | 75% | N/A | 25% |
| Apostolopoulos37 | Public database & Literature | Radiology images | 1427 | 10-fold cross-validation | 90% | N/A | 10% |
| Bai95 | Clinical setting | Radiology images, clinical data, and laboratory data | 133 | 5-fold cross-validation | 80% | N/A | 20% |
| Barstugan21 | Public database | Radiology images | 150 | 2-fold, 5-fold and 10-fold cross-validations | 50%, 80%, 90% | N/A | 10%,20%, 50% |
| Beck61 | Public database | Biological data | N/A | N/A | N/A | N/A | N/A |
| Bukhari38 | Public database | Radiology images | 278 | Training-test split | 80% | N/A | 20% |
| Chen J22 | Clinical setting | Radiology images, clinical data, and laboratory data | 24907 | Training-test split | 4% | 40% | 56% |
| Chen X89 | Public database | Radiology images | 110 | 10-fold cross-validation | 90% | N/A | 10% |
| Chowdhury39 | Public database & Literature | Radiology images | 8482 | 5-fold cross-validation | 96% | N/A | 4% |
| Dandekar73 | Government sources | Epidemiological Data | N/A | N/A | N/A | N/A | N/A |
| DeCapprio96 | Government sources | Clinical data & Demographic data | 1851519 | Training-test split | 80% | N/A | 20% |
| Dutta74 | Public database | Epidemiological Data | N/A | N/A | N/A | N/A | N/A |
| Fast68 | Public database | Biological data | N/A | N/A | N/A | N/A | N/A |
| Feng49 | Clinical setting | Clinical data, demographic data, laboratory data | 132 | 10-fold cross-validation | 80% | N/A | 20% |
| Fong-a75 | Government sources | Epidemiological Data | N/A | N/A | N/A | N/A | N/A |
| Fong-b76 | Government sources | Epidemiological Data | N/A | N/A | N/A | N/A | N/A |
| Fu23 | Clinical setting | Radiology images | 89628 | External validation | 50% | 17% | 33% |
| Gaal90 | Public database | Radiology images | 1047 | Training-test split | N/A | N/A | N/A |
| Gao52 | Public database | Biological data | N/A | N/A | N/A | N/A | N/A |
| Ge53 | Public database | Biological data | N/A | N/A | N/A | N/A | N/A |
| Ghoshal40 | Public database | Radiology images | 5941 | Training-test split | 80% | N/A | 20% |
| Gong97 | Clinical setting | Clinical data, demographic data, laboratory data | 372 | Training-test split | 51% | 44% | 5% |
| Gozes24 | Public database & Clinical setting | Radiology images | 207 | External validation | 24% | N/A | 76% |
| Guo86 | Public database | Biological data | N/A | N/A | N/A | N/A | N/A |
| Hemdan41 | Public database | Radiology images | 50 | Training-test split | 80% | N/A | 20% |
| Heo69 | Public database | Biological data | N/A | N/A | N/A | N/A | N/A |
| Hofmarcher54 | Public database | Biological data | N/A | N/A | N/A | N/A | N/A |
| Hu-a62 | Public database | Biological data | N/A | N/A | N/A | N/A | N/A |
| Hu-b77 | Public database | Epidemiological Data | N/A | N/A | N/A | N/A | N/A |
| Hu-c78 | Public database | Epidemiological Data | N/A | N/A | N/A | N/A | N/A |
| Huang C79 | Public database | Epidemiological Data | N/A | External validation | N/A | N/A | N/A |
| Huang L91 | Clinical setting | Radiology images | 842 | Training-test split | 91% | N/A | 9% |
| Jin C25 | Public database & Clinical setting | Radiology images | 1671 | External validation | 19% | 6% | 75% |
| Jin S26 | Clinical setting | Radiology images | 1418 | Training-test split | 80% | N/A | 20% |
| Kumar80 | Public database | Epidemiological Data | N/A | N/A | N/A | N/A | N/A |
| Li L27 | Clinical setting | Radiology images | 4356 | Training-test split | 90% | N/A | 10% |
| Li M81 | Public database & Government sources | Epidemiological Data | N/A | N/A | N/A | N/A | N/A |
| Lopez-Rincon50 | Public database & Government sources | Biological data | 553 | 10-fold cross-validation | 80% | 10% | 10% |
| Magar55 | Public database | Biological data | 1933 | 5-fold cross-validation | 80% | N/A | 20% |
| Marini82 | Public database | Epidemiological Data | N/A | N/A | N/A | N/A | N/A |
| Meng48 | Clinical setting | Clinical data & laboratory data | 620 | External validation | 70% | N/A | 30% |
| Mizumoto83 | Government sources | Laboratory data | 634 | N/A | N/A | N/A | N/A |
| Narin42 | Public database | Radiology images | 100 | 5-fold cross-validation | 80% | N/A | 20% |
| Ong65 | Public database | Biological data | N/A | 5-fold cross-validation | N/A | N/A | N/A |
| Ozturk43 | Public database | Radiology images | 1125 | 5-fold cross-validation | 80% | N/A | 20% |
| Pandey102 | Public database & News websites | Guidelines & news | N/A | N/A | N/A | N/A | N/A |
| Patankar56 | Public database | Biological data | 310000 | Training-test split | 75% | N/A | 25% |
| Pirouz84 | Public database | Epidemiological Data | 64786 | Training-test split | 75% | N/A | 25% |
| Pourhomayoun98 | Literature, Government sources, News websites | Clinical data & Demographic data | 117000 | 10-fold cross-validation | 90% | N/A | 10% |
| Qi101 | Clinical setting | Radiology images | 31 | 5-fold cross-validation | 80% | N/A | 20% |
| Qiang87 | Government sources | Biological data | 2666 | 10-fold cross-validation | 90% | N/A | 10% |
| Qiao70 | Public database | Biological data | N/A | N/A | N/A | N/A | N/A |
| Rahman66 | Public database | Biological data | 250 | N/A | N/A | N/A | N/A |
| Randhawa88 | Public database | Biological data | N/A | 10-fold cross-validation | 90% | N/A | 10% |
| Saçar demirci71 | Public database | Biological data | N/A | N/A | N/A | N/A | N/A |
| Sarkar B67 | Public database | Biological data | N/A | N/A | N/A | N/A | N/A |
| Sarkar J100 | Public database | Clinical data | 430 | Training-test split | 70% | N/A | 30% |
| Sethy44 | Public database | Radiology images | 183 | Training-test split | 60% | 20% | 20% |
| Shan92 | Clinical setting | Radiology images | 549 | External validation | 45% | N/A | 55% |
| Shi28 | Clinical setting | Radiology images | 2685 | 5-fold cross-validation | 80% | N/A | 20% |
| Tang B57 | Public database | Biological data | N/A | N/A | N/A | N/A | N/A |
| Tang Z93 | Clinical setting | Radiology images | 176 | 3-fold cross-validation | 65% | N/A | 35% |
| Tiwari85 | Public database | Epidemiological Data | N/A | Training-test split | 70% | N/A | 30% |
| Ton58 | Public database & Literature | Biological data | 3000000 | N/A | N/A | N/A | N/A |
| Ucar45 | Public database | Radiology images | 5949 | Training-test split | 80% | 10% | 10% |
| Wang L46 | Public database | Radiology images | N/A | Training-test split | N/A | N/A | N/A |
| Wang Shuai29 | Clinical setting | Radiology images | 453 | External validation | 52% | N/A | 48% |
| Wang Shuo30 | Clinical setting | Radiology images | 1266 | External validation | 56% | N/A | 44% |
| Wang Y51 | Participants | Radiology images | 120605 | External validation | 99.50% | N/A | 0.50% |
| Wang Z63 | Literature | Clinical data | N/A | 10-fold cross-validation | 90% | N/A | 10% |
| Xu31 | Clinical setting | Radiology images | 618 | Training-test split | 85% | N/A | 15% |
| Yan99 | Clinical setting | Clinical data, demographic data, laboratory data, epidemiological data | 404 | External validation | 66% | 27% | 7% |
| Ying32 | Clinical setting | Radiology images | 1990 | Training-test split | 60% | 10% | 30% |
| Yu94 | Clinical setting | Clinical data, laboratory data, epidemiological data | 105 | N/A | N/A | N/A | N/A |
| Zhang H59 | Public database | Biological data | N/A | N/A | N/A | N/A | N/A |
| Zhang J47 | Public database | Radiology images | 1531 | Training-test split | 93% | N/A | 7% |
| Zhao33 | Public database & Literature | Radiology images | 471 | Training-test split | 60% | 15% | 15% |
| Zhavoronkov60 | Public database & Clinical setting | Biological data | N/A | N/A | N/A | N/A | N/A |
| Zheng34 | Clinical setting | Radiology images | 630 | Training-test split | 79% | N/A | 21% |
| Zhou35 | Clinical setting | Radiology images | 4474 | Training-test split & External validation | 63% | 7% | 30% |
| Abbreviations | N/A: not available |  |  |  |  |  |  |