Experimental Methods with Parameters

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | Classifier | Cross-validation function | Cross-validation Parameter | Parameter |
| WBCD | NB | KFold | n\_splits=10, shuffle=True,  random\_state=101 | default |
| LR | RepeatedKFold | n\_splits=8,  random\_state=51 | default |
| DT | StratifiedKFold | n\_splits=5, shuffle=True,  random\_state=101 | criterion="entropy",  max\_depth=3 |
| SVM | StratifiedKFold | n\_splits=8, shuffle=True,  random\_state=101 | kernel='linear' |
| LDA | KFold | n\_splits=6, shuffle=True,  random\_state=101 | default |
| K-NN | StratifiedKFold | n\_splits=5, shuffle=True,  random\_state=51 | n\_neighbors=3 |
| AD | KFold | n\_splits=5, shuffle=True  random\_state=101 | default |
| RF | RepeatedKFold | n\_splits=10,  random\_state=101 | default |
| VC | RepeatedStratifiedKFold | n\_splits=10,  random\_state=101 | estimators= [  AdaBoostClassifier (),  Logistic Regression (),  RandomForestClassifier(),  SVC ()] |
| SGD | StratifiedKFold | n\_splits=10, shuffle=True,  random\_state=51 | max\_iter=500, random\_state=51 |
| BC | KFold | n\_splits=10, shuffle=True,  random\_state=101 | base\_estimator = (RandomForestClassifier  (), n\_estimators = 100) |
| GB | KFold | n\_splits=10, shuffle=True,  random\_state=51 | default |
| WBC | NB | KFold | n\_splits=10, shuffle =True,  random\_state=101 | default |
| LR | KFold | n\_splits=10, shuffle=True,  random\_state=101 | default |
| DT | StratifiedKFold | n\_splits=10, shuffle=True,  random\_state=101 | criterion="entropy",  max\_depth=3 |
| SVM | KFold | n\_splits=10, shuffle=True,  random\_state=101 | default |
| LDA | KFold | n\_splits=10, shuffle=True,  random\_state=101 | default |
| K-NN | KFold | n\_splits=10, shuffle=True,  random\_state=101 | default |
| AD | KFold | n\_splits=10, shuffle=True,  random\_state=101 | default |
| RF | KFold | n\_splits=10, shuffle=True,  random\_state=101 | default |
| VC | KFold | n\_splits=10, shuffle=True,  random\_state=101 | estimators= [  AdaBoostClassifier (),  Logistic Regression (),  RandomForestClassifier(),  SVC()] |
| SGD | RepeatedStratifiedKFold | n\_splits=10,  random\_state=101 | max\_iter=100 |
| BC | KFold | n\_splits=10, shuffle=True,  random\_state=101 | base\_estimator =  (RandomForestClassifier  (), n\_estimators = 100) |
| GB | KFold | n\_splits=10, shuffle=True,  random\_state=101 | default |

On the two datasets, we conducted twelve classifiers in this study. The two datasets have the different number of features and also their correlation is different. However, these two datasets infected the classifier in different ways, so we used different parameters to improve performance. Furthermore, for the imbalance of the two datasets, we used different cross-validation functions with different k values based on the classifier's mechanism. This study's investigation is represented in Table II.