In this section, we have discussed about model performance. The discussion divides into two parts: The performance of the model without cross-validation and the performance of the model with cross-validation.

Using two breast cancer datasets, we conducted a comparison study of twelve machine learning algorithms. To train the model, divide the dataset into two parts: training and testing. On the WBC dataset, 512 samples (90%) were used for training and 57 samples (10%) were used for testing, while 629 samples (90%) were used for training and 70 samples (10%) were used for testing. Furthermore, the performance of all classifiers has been evaluated in terms of accuracy, precision, F1-measures, specificity, sensitivity, False Discovery Rate (FDR), False Omission Rate (FOR) (FOR).

**The performance of the model without cross-validation**

**Performance of Training Phase**

The results of our breast cancer detection experiments on the WBCD and WBC datasets are shown below. Table-2 displays the False Negative value of twelve algorithms. False negative value is a critical issue in the detection of breast cancer.

Table-2: False Negative Value

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Dataset | NB | LR | DT | SVM | LDA | KNN | AB | RF | VC | SGD | BC | GB |
| WBCD | 19 | 5 | 16 | 5 | 16 | 10 | 0 | 0 | 5 | 3 | 3 | 0 |
| WBC | 6 | 11 | 1 | 4 | 18 | 6 | 5 | 0 | 5 | 6 | 1 | 0 |

Table-2 clearly shows that three algorithms, AB, RF, and GB on the WBCD dataset, and two algorithms, RF and GB on the WBC dataset, do not produce any False Negative values, indicating that those algorithms correctly classified every sample. The reason for giving those algorithms a False Negative value of zero is that they are ensemble and boosting algorithms. Furthermore, we thoroughly investigate the model performance in light of the performance metrics. The model performance is shown in table-3.

Table-3: Model Performance Metrics in the training phase on the WBCD Dataset

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Parameters | NB | LR | DT | SVM | LDA | KNN | AB | RF | VC | SGD | BC | GB |
| Precision | 0.96 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 1.0 | 1.0 | 1.0 | 0.99 | 1.0 | 1.0 |
| Recall or Sensitivity | 0.94 | 0.98 | 0.95 | 0.98 | 0.95 | 0.96 | 1.0 | 1.0 | 0.98 | 0.99 | 0.99 | 1.0 |
| Specificity | 0.94 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 1.0 | 1.0 | 1.0 | 0.99 | 1.0 | 1.0 |
| F1-Score | 0.95 | 0.99 | 0.97 | 0.99 | 0.97 | 0.98 | 1.0 | 1.0 | 0.99 | 0.99 | 0.99 | 1.0 |
| False Discovery Rate (FDR) | 0.038 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.0 | 0.0 | 0.0 | 0.006 | 0.0 | 0.0 |
| False Omission Rate (FOR) | 0.095 | 0.025 | 0.08 | 0.025 | 0.08 | 0.05 | 0.0 | 0.0 | 0.025 | 0.015 | 0.015 | 0.0 |

The table- shows the all-performance metrics on the WBCD dataset. Here, AB, RF and GB provides best result than others model. In addition, DT, SVM, LDA, KNN and BC also generates less false discovery rate with high false omission rate. Further, f1-score, specificity, and precision are almost nearly 1. Also, we investigate the model performance metrics on the WBC dataset.

Table-4: Model Performance Metrics on the WBC Dataset

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Parameters | NB | LR | DT | SVM | LDA | KNN | AB | RF | VC | SGD | BC | GB |
| Precision | 0.949 | 0.974 | 0.945 | 0.969 | 0.978 | 0.974 | 0.983 | 1.0 | 0.986 | 0.964 | 0.983 | 1.0 |
| Recall or Sensitivity | 0.985 | 0.974 | 0.997 | 0.990 | 0.958 | 0.985 | 0.988 | 1.0 | 0.988 | 0.993 | 0.998 | 1.0 |
| Specificity | 0.907 | 0.948 | 0.902 | 0.941 | 0.956 | 0.949 | 0.967 | 1.0 | 0.972 | 0.933 | 0.968 | 1.0 |
| F1-Score | 0.967 | 0.974 | 0.970 | 0.979 | 0.968 | 0.979 | 0.986 | 1.0 | 0.987 | 0.978 | 0.990 | 1.0 |
| False Discovery Rate (FDR) | 0.050 | 0.026 | 0.055 | 0.031 | 0.022 | 0.026 | 0.017 | 0.0 | 0.014 | 0.035 | 0.017 | 0.0 |
| False Omission Rate (FOR) | 0.028 | 0.052 | 0.005 | 0.019 | 0.085 | 0.028 | 0.024 | 0.0 | 0.023 | 0.014 | 0.004 | 0.0 |

The table-4 shows the all-performance metrics on the WBC dataset. From this table it is a clear that RF and GB model is the best model according to the whole metrics. Moreover, Recall or Sensitivity and F1-Score are also high of SVM, DT, SGD and BC models.

Table-5: Accuracy on the WBCD Dataset

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Parameters | NB | LR | DT | SVM | LDA | KNN | AB | RF | VC | SGD | BC | GB |
| Accuracy | 93.94 | 98.83 | 96.68 | 98.83 | 96.68 | 97.85 | 100 | 100 | 99.02 | 99.02 | 99.41 | 100 |

Table-5 displays the accuracy of various classifiers on the WBCD dataset. In this case, most classifiers provide accuracy greater than 95% accept NB. Three classifiers, AB, RF, and GB, among them, provide 100% accuracy. Furthermore, VC, SGD and BC classifiers achieved the accuracy 99.02%, 99.02% and 99.41% respectively.

Table-6: Accuracy on the WBC Dataset

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Parameters | NB | LR | DT | SVM | LDA | KNN | AB | RF | VC | SGD | BC | GB |
| Accuracy | 95.707 | 96.502 | 96.184 | 97.297 | 95.707 | 97.297 | 98.092 | 100.0 | 98.251 | 97.138 | 98.728 | 100.0 |

In addition, table-6 displays the accuracy of different classifiers on the WBC dataset. Also, almost every classifier here provides 95% above accuracy, but the range of accuracy is less than in table-4. Also, AB, VC and BC classifiers obtained the 98.092%, 98.251% and 98.728% accuracy.

**Performance of Testing Phase**

Table-7: False Negative Value

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Dataset | NB | LR | DT | SVM | LDA | KNN | AB | RF | VC | SGD | BC | GB |
| WBCD | 0 | 0 | 0 | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| WBC | 0 | 0 | 1 | 0 | 1 | 2 | 0 | 0 | 0 | 1 | 0 | 1 |

From both datasets, it can be observed without a doubt from Table 7 that the seven algorithms such as NB, LR, SVM, AB, RF, VC and BC do not provide any False Negative values, indicating that accurately identified each sample. Moreover, the three algorithms (DT, SGD and GB) do not provide any False Negative values on the WBCD dataset.

Table-8: Model Performance Metrics on the WBCD Dataset

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Parameters | NB | LR | DT | SVM | LDA | KNN | AB | RF | VC | SGD | BC | GB |
| Precision | 0.95 | 0.93 | 0.88 | 0.93 | 0.95 | 0.98 | 0.88 | 0.95 | 0.93 | 0.84 | 0.93 | 0.93 |
| Recall or Sensitivity | 1.0 | 1.0 | 1.0 | 1.0 | 0.95 | 0.98 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| Specificity | 0.87 | 0.82 | 0.73 | 0.82 | 0.86 | 0.93 | 0.74 | 0.88 | 0.82 | 0.67 | 0.82 | 0.82 |
| F1-Score | 0.97 | 0.96 | 0.93 | 0.96 | 0.95 | 0.98 | 0.94 | 0.98 | 0.96 | 0.91 | 0.96 | 0.96 |
| False Discovery Rate (FDR) | 0.04 | 0.06 | 0.11 | 0.06 | 0.046 | 0.02 | 0.116 | 0.04 | 0.069 | 0.163 | 0.069 | 0.069 |
| False Omission Rate (FOR) | 0.0 | 0.0 | 0.0 | 0.0 | 0.14 | 0.07 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |

In the Table 8, several classifiers such as NB, LR, LDA, KNN and RF shows better performance. Hence, On the other performance parameters of classifiers are standard.

Table-9: Model Performance Metrics on the WBC Dataset

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Parameters | NB | LR | DT | SVM | LDA | KNN | AB | RF | VC | SGD | BC | GB |
| Precision | 0.976 | 1.0 | 0.976 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 0.976 |
| Recall or Sensitivity | 1.0 | 1.0 | 0.976 | 1.0 | 0.976 | ­­­­­0.953 | 1.0 | 1.0 | 1.0 | 0.976 | 1.0 | 0.976 |
| Specificity | 0.967 | 1.0 | 0.966 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 0.966 |
| F1-Score | 0.988 | 1.0 | 0.976 | 1.0 | 0.988 | 0.976 | 1.0 | 1.0 | 1.0 | 0.988 | 1.0 | 0.976 |
| False Discovery Rate (FDR) | 0.024 | 0.0 | 0.024 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.024 |
| False Omission Rate (FOR) | 0.0 | 0.0 | 0.034 | 0.0 | 0.034 | 0.069 | 0.0 | 0.0 | 0.0 | 0.034 | 0.0 | 0.034 |

In overall, the outperformance of all the classifiers is incredibly high. Additionally, NB, DT, and GB classifiers provide well-standard.

Table-10: Accuracy on the WBCD Dataset

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Parameters | NB | LR | DT | SVM | LDA | KNN | AB | RF | VC | SGD | BC | GB |
| Accuracy | 96.49 | 94.74 | 91.23 | 96.49 | 92.98 | 96.49 | 91.23 | 96.49 | 94.74 | 87.72 | 94.74 | 94.74 |

From table 10, we can observe that the performance of each classifier varies, but NB, KNN and RF consistently achieves the highest accuracy on the testing set (96.49%).

Table-11: Accuracy on the WBC Dataset

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Parameters | NB | LR | DT | SVM | LDA | KNN | AB | RF | VC | SGD | BC | GB |
| Accuracy | 98.571 | 100.0 | 97.143 | 100.0 | 98.571 | 97.143 | 100.0 | 100.0 | 100.0 | 98.571 | 100.0 | 97.143 |

Owing to cross validation on the WBC dataset, boosts the accurateness of model where are appropriate accuracy achieves in LR (98.6%), LDA (98.57%), KNN (98.97%), AB (98.57%), RF (98.7%), SGD (98.57%), BC (98.6%) and GB (98.57%).

**The performance of the model with cross-validation**

**Training Accuracy**

Since the datasets are imbalance for that some classifiers provides high accuracy. Because the dataset distribution is not uniform, the probabilistic model produced good results for certain test sets. Cross-validation is the most commonly used technique to solve this problem.

Table-12: Accuracy on the WBCD Dataset

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Parameters | NB | LR | DT | SVM | LDA | KNN | AB | RF | VC | SGD | BC | GB |
| Cross validation | 93.15 | 97.91 | 93.75 | 97.85 | 96.49 | 96.68 | 96.29 | 96.03 | 97.93 | 96.88 | 95.51 | 96.29 |

The table-12 presents the cross-validation accuracy of twelve classifiers in the training phase on the WBCD dataset. Almost every classifier's cross-validation accuracy is lower than its original accuracy. It indicates that the imbalanced dataset had an effect on the classifier.

Table-13: Accuracy on the WBC Dataset

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Parameters | NB | LR | DT | SVM | LDA | KNN | AB | RF | VC | SGD | BC | GB |
| Cross validation | 95.691 | 96.503 | 95.233 | 96.503 | 95.230 | 97.140 | 95.709 | 97.140 | 96.662 | 95.947 | 97.299 | 96.188 |

Table-13 also displays the cross-validation accuracy of twelve classifiers during the training phase on the WBC dataset. In this case, two classifiers, NB and LR, outperform the original accuracy. Furthermore, the other two classifiers, LDA and KNN, provide nearly the same accuracy as the original accuracy.

**Testing Accuracy**

Table-14: Accuracy on the WBCD Dataset

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Parameters | NB | LR | DT | SVM | LDA | KNN | AB | RF | VC | SGD | BC | GB |
| Cross validation | 95.0 | 96.67 | 91.67 | 93.07 | 90.0 | 93.07 | 91.48 | 96.67 | 94.67 | 98.33 | 94.6 | 91.67 |

The cross-validation accuracy of twelve classifiers during testing on the WBCD dataset is shown in Table 14. The results of LR, DT, AB, RF, and SGD are superior to the original results, as can be seen from the table. For this, we argued that these classifiers are not overfit.

Table-15: Accuracy on the WBC Dataset

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Parameters | NB | LR | DT | SVM | LDA | KNN | AB | RF | VC | SGD | BC | GB |
| Cross validation | 95.714 | 98.611 | 95.714 | 98.00 | 98.733 | 98.872 | 98.571 | 98.715 | 98.733 | 98.571 | 98.611 | 98.571 |

Likewise, the table-15 represents the cross-validation accuracy of twelve classifiers in the testing phase on the WBC dataset. Also, here three classifiers LDA, KNN and GB outperform original result. In addition, others classifiers also perform well. We see that LR, AB, RF, VC, SGD, and BC classifiers provide the 98.5% up accuracy. So, we conclude that these classifiers perform well on the test dataset.