|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| References | Sensitivity | Specificity | PPV | NPV | AUC | Accuracy | |
| Exarchos KP, et al. 2012.34  RF | 0.8 (± 14.56) | 0.7 (± 23.90) | ND | ND | 0.8 (± 0.10) | 0.8 (± 8.96) | |
| Exarchos KP, et al. 2012.35 | Multiple clinical, radiological, and genomic data combined with different results and metrics. | | | | | | |
| Chang SW, et al. 2013.36 | ND | ND | ND | ND | 0.90 | | 0.93 |
| Alabi RO, et al. 2019.37 | 0.71 | 0.98 | 0.97 | 0.84 | 0.97 | | 0.88 |
| Alabi RO, et al. 2020.38\*  SVM  NB  BDT  DF | 0.84  0.84  0.79  0.79 | 0.60  0.63  0.83  0.78 | 0.50  0.52  0.76  0.63 | 0.89  0.89  0.89  0.89 | ND  ND  ND  ND | | 68  70  81  78 |
| Bur AM, et al. 2019.39  SVM  Gradient Boosting  LR  DF | 0.83  0.91  0.91  0.91 | 0.55  0.52  0.62  057 | ND  ND  ND  ND | ND  ND  ND  ND | 0.776  0.798  0.821  0.840 | | ND  ND  ND  ND |
| Mermod M, et al. 2019.40  RF  SVM  Lasso | 0.8 (CI: 0.6-0.9)  0.8 (CI: 0.6-1)  0.9 (CI: 0.7-1) | 0.9 (CI: 0.8-0.9)  0.7 (CI: 0.6-0.8)  0.5 (CI: 0.4-0.6) | 0.6 (CI: 0.4-0.8)  0.4 (CI: 0.3-0.6)  0.3 (CI: 0.2-0.4) | 0.95 (CI: 0.9-1)  1 (CI: 0.9-1)  1 (CI: 0.8-1) | 0.85 (CI: 0.8-0.9)  0.8 (CI: 0.8-1)  0.8 (CI: 0.7-0.9) | | 0.9 (CI: 0.8-0.9)  0.7 (CI: 0.7-0.8)  0.6 (CI: 0.5-0.7) |
| Karadaghy OA, et al. 2019.41  DF  2-Class DF | 0.68  0.52 | ND  ND | 0.71  0.69 | ND  ND | 0.80 (CI: 0.79-0.81)  0.68 (CI: 0.67-0.70) | | 0.71  0.65 |

**Supplementary Table 1**. Metrics from studies included. \*Performance of the algorithms with external cases. Abbreviations: Abbreviations: DBN = Dynamic Bayesian Network; SVM = Support Vector Machine; LR = Logistic Regression; DF = Decision Forest; RF = Random forest; GB = gradient boosting; BDT = Boosted Decision Tree; NB= Naive Boyes; DT = Decision tree; CI = Confidence interval.

|  |  |  |  |
| --- | --- | --- | --- |
| Study | Missing data | Addressed | Mechanism. |
| Exarchos KP, et al.34 | Yes | Yes | Features with high percentage (>90%) of missing values are omitted from their analysis.  Features with less percentage of missing values are imputed with the modes and means.  Enrolled patients were unevenly distributed in the classes of relapsers and non-relapsers, resulting in considerable class imbalance. For this purpose, authors employ the Synthetic Minority Oversampling Technique. |
| Exarchos KP, et al.35 | Yes | Yes | Features with high percentage (>90%) of missing values are  omitted from our analysis.  Features with less percentage of missing values are imputed with the modes and means. |
| Chang SW, et al.36 | Yes | Yes | Authors employed the feature selection methods in their dataset to choose the most optimum feature subsets based on the correlations of the input and output variables. |
| Alabi RO, et al.37 | Yes | No | Authors do not consider missing data and used only a complete cases analysis. |
| Alabi RO, et al.38 | Yes | Yes | Class imbalance was addressed employing the Synthetic Minority Oversampling Technique. |
| Bur AM, et al.39 | Yes | Yes | Missing data for continuous variables were handled using median imputation.  Missing categorical data was assigned a value of unknown. |
| Mermod M, et al.40 | Yes | Yes | Given the difference of T classification the ratio of pN+ and pN0. Authors compensated the training cohort by a combination of under sampling and oversampling, implementing the Random Over-Sampling Examples strategy |
| Karadaghy OA. 2019.41 | Yes | Yes | Missing data for covariates of interest were explored and categorized  (missing completely at random, missing at random and missing not at random).  Variables determined to be missing at random were handled using single value imputation of median values. No data imputation was used for variables determined to be missing not at random.  No data imputation was used for variables with missing information  greater than 40%. |

**Supplementary table 2**. Missing data information of studies included.