Vol.66: e23230002, 2023
https://doi.org/10.1590/1678-4324-2023230002
ISSN 1678-4324 Online Edition

*Article – Engineering, Technology and Techniques*

# Classification of 1p/19q Status in Low-Grade Gliomas: Experiments with Radiomic Features and Ensemble-Based Machine Learning Methods

**Tony Alexandre Medeiros**[1,2*]
https://orcid.org/0000-0001-5895-8439

**Francisco Assis de Oliveira Nascimento**[1]
https://orcid.org/0000-0002-8217-1983

**Raimundo Guimarães Saraiva Junior**[1,3]
https://orcid.org/0000-0002-4083-9404

**João Luiz Azevedo de Carvalho**[1]
https://orcid.org/0000-0002-6485-6380

**Guilherme de Souza e Cassia**[4]
https://orcid.org/0000-0001-9029-4679

[1]Universidade de Brasília, Departamento de Engenharia Elétrica, Brasília, Distrito Federal, Brasil; [2]Instituto Federal Goiano, Departamento de Informática, Cristalina, Goiás, Brasil; [3]Instituto Federal do Ceará, Departamento de Eletrônica, Limoeiro do Norte, Ceará, Brasil; [4]Rede D'Or São Luiz, Hospital Santa Luzia, Brasília, Distrito Federal, Brasil.

Editor-in-Chief: Bill Jorge Costa
Associate Editor: Fabio Alessandro Guerra

Received: 09-Jan-2023; Accepted: 07-Mar-2023

*Correspondence: tony.medeiros@ifgoiano.edu.br; Tel.: +55-61-98663-9824 (T.A.M.)

**HIGHLIGHTS**

- We investigate machine learning assessment of 1p/19q status in low grade gliomas.
- Experiments were performed with different sets of radiomic features.
- Several classifiers were evaluated, including various ensemble methods.
- Best results achieved using a bagging estimator with texture-based radiomic features.

**Abstract:** Gliomas comprise the vast majority of all malignant brain tumors. Low-grade glioma patients with combined whole-arm losses of 1p and 19q chromosomes were shown to have significantly better overall survival rates compared to non-deleted patients. This work evaluates several approaches for assessment of 1p/19q status from T2-weighted magnetic resonance images, using radiomics and machine learning. Experiments were performed using images from a public database (102 codeleted, 57 non-deleted). We experimented with sets of 68 and 100 radiomic features, and with several classifiers, including support vector machine, k-nearest neighbors, stochastic gradient descent, logistic regression, decision tree, Gaussian naive Bayes, and linear discriminant analysis. We also experimented with several ensemble-based methods, including four boosting-based classifiers, random forest, extra-trees, and bagging. The performance of these methods was compared using various metrics. Our best results were achieved using a bagging ensemble estimator based on the decision tree classifier, using only texture-based radiomics features. Compared to other works that used the same database, this approach provided higher sensitivity. It also achieved higher sensitivity than that provided by neurosurgeons and neuroradiologists analyzing the same images. We also show that including radiomic features associated with first order statistics and shape does not improve the performance of the classifiers, and in many cases worsens it. The molecular assessment of brain tumors

through biopsies is an invasive procedure, and is subject to sampling errors. Therefore, the techniques presented in this work have strong potential for aiding in better clinical, surgical, and therapeutic decision-making.

**Keywords:** low-grade glioma; 1p/19q codeletion; radiomics; machine learning; ensemble methods.

## INTRODUCTION

Gliomas comprise the vast majority of all malignant brain tumors. The World Health Organization recommends the use of an integrative classification of tumors, considering histopathological and molecular characteristics, thus allowing for better clinical management [1]. For example, oligodendrogliomas are diagnosed by the identification of the combined whole-arm losses of 1p and 19q chromosomes (1p/19q codeletion). Oligodendrogliomas have better prognosis than astrocytomas and glioblastomas; thus, the identification of 1p/19q status (codeleted or non-deleted) is a very important step for evaluating outcomes in patients with gliomas [2,3]. Low-grade glioma (LGG) patients with codeleted 1p/19q status were shown to have significantly better overall survival rates compared to non-deleted patients [4]. Thus, knowledge of 1p/19q is of fundamental importance in making better clinical, surgical, and therapeutic decisions.

The molecular assessment of brain tumors is made through biopsies, which are invasive procedures that are subject to sampling errors, involve surgical risks [5]. One non-invasive alternative for assessment of 1p/19q status involves the use of magnetic resonance imaging (MRI), radiomics, and machine learning. MRI is a non-invasive medical imaging method, as it does not use ionizing radiation. MRI provides superior soft tissue contrast, and therefore it is the gold standard imaging method for the diagnosis and treatment management of brain tumors [6,7]. Radiomics is a tool for extracting quantitative attributes from medical images, allowing the mining of non-visual information, which may correlate to histopathological and genetic characteristics of the tumor, enabling molecular diagnosis, assessment of tumor grade and prognosis, as well as assisting in clinical decision-making, guiding treatments and pre-surgical mapping [8,9]. Machine learning provides techniques so that a computational model can be created from data (input) with a known response (output) [10].

Many research groups have been working on machine learning methods for assessment of 1p/19q status [11-20]. The works of those groups has shown that machine learning and deep learning have strong potential for diagnosis of 1p/19q codeletion. Several authors have demonstrated machine learning methods which perform better than radiologists diagnosing brain tumors in MRI images [13]. However, there are several machine learning tools yet unexplored for assessment of 1p/19q status, some of which may provide even better performance for this application.

This work evaluates several approaches for the non-invasive evaluation of the 1p/19q chromosome codeletion status from T2-weighted magnetic resonance images, using radiomics and machine learning. The evaluated methods use radiomics feature extraction, feature scaling, and classification based on various supervised models. Our best results were achieved using a bagging ensemble estimator based on the decision tree classifier, using only texture-based radiomics features. Compared to other works that used the same database, this approach provided higher sensitivity. It also achieved higher sensitivity than that provided by neurosurgeons and neuroradiologists analyzing the same images. We also show that including radiomic features associated with first order statistics and shape does not improve the performance of the classifiers, and in many cases worsens it.

## METHODS

### Dataset

All experiments were performed using the LGG-1p19qDeletion dataset from The Cancer Imaging Archive [20,21]. This dataset consists of MRI images acquired at 1.5 T or 3 T, including post-contrast T1- and T2-weighted images (1-mm and 3-mm slice thickness, respectively). However, we only used the T2-weighted images, because these better highlight fluid and edema compared to other sequences, and therefore are very useful in detailing various pathological processes, including for determining the extent of infiltration of a brain tumor. The dataset contains pre-operative images from 159 patients, all with 1p/19q status proven through biopsy (102 patients with codeleted LGG, and 57 patients with non-deleted LGG). Representative images are shown in Figure 1. The dataset also includes semi-automatically segmented regions-of-interest (ROI) for all images. Each ROI completely encloses a tumor and its boundaries [20], as illustrated in Figure 2.

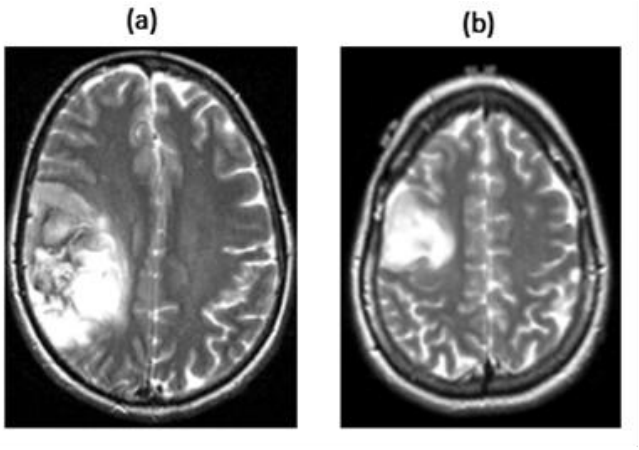


**Figure 1**. Representative images from the LGG-1p19q. Deletion database: (a) large heterogeneous infiltrative lesion in the right frontoparietal region with areas of low signal intensity, suggesting hemorrhage or calcification, diagnosed as oligodendroglioma (1p/19q codeleted tumor); and (b) homogeneous infiltrative lesion in the right frontal lobe with high signal intensity, diagnosed as astrocytoma (1p/19q non-codeleted tumor).

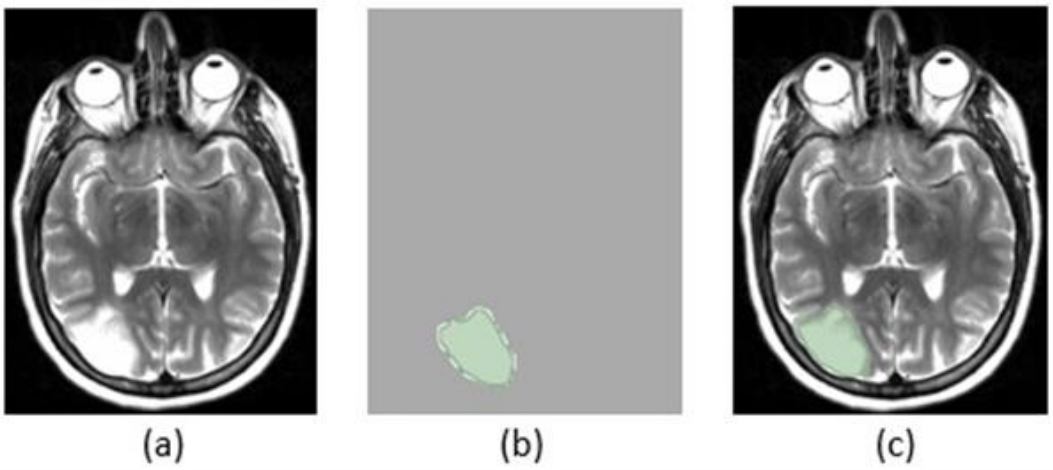


**Figure 2.** Example of a segmented tumor from the LGG-1p19qDeletion database: (a) image containing right occipital tumor with homogeneous high signal intensity, diagnosed as oligodendroglioma (1p/19q codeleted); (b) semi-automatically segmented region-of-interest; and (c) region-of-interest overlayed over image.

## Feature extraction

The extraction of quantitative features from the tumors was performed using the PyRadiomics platform version 2.0 [22]. PyRadiomics is a flexible open-source plat- form, which allows extracting a large number of radiomic features from medical images.

The ROIs in the LGG-1p19qDeletion database are provided in the Neuroimaging Informatics Technology Initiative (NIfTI) format, which is not supported in PyRadiomics. Thus, these were converted to the Nearly Raw Raster Data (NRRD) format using the 3DSlicer software tool [23].

We extracted 100 features from each ROI, with no filters enabled. The features in PyRadiomics are subdivided into classes, including: first order statistics (first order); shape-based (shape); gray level cooccurence matrix (GLCM); gray level run length matrix (GLRLM); gray level size zone matrix (GLSZM); and gray level dependence matrix (GLDM). A summary of the features we extracted is presented in Table 1. Note that while first order statistics features describe the distribution of voxel intensities within the ROI, and shape-based features describe size and shape of the ROI, the remaining classes of features describe texture characteristics [22,24]. We experimented with sets of 68 and 100 features, as will be discussed later. The 68-feature set uses only texture-based features.

**Table 1.** Summary of extracted radiomic features.

| Class Name | Number of Features | Class Description |
|---|---|---|
| first order | 18∗ | Describe the distribution of voxel intensities within the ROI |
| shape | 14∗ | Describe the three-dimensional size and shape of the ROI |
| GLCM | 22 | Quantifies the cooccurence of voxel intensities within the ROI |
| GLRLM | 16 | Quantifies the occurrences of consecutive voxels within the ROI that have the same gray level |
| GLSZM | 16 | Quantifies the number of connected voxels that share the same gray level intensity in the ROI |
| GLDM | 14 | Quantifies gray level dependencies in the ROI |

∗ We experimented with sets of 68 and 100 features. First order statistics and shape-based features were not used when working with only 68 features.

### Train-test data split

After feature extraction, the dataset with 159 instances (57 non-deleted and 102 codeleted) was randomly divided into two subsets: 90% for training (143 instances) and 10% for testing (16 instances).

### Feature normalization

Different radiomic features come in different orders of magnitude. Thus, normalization was used so that all features display similar ranges of values. This helps the classification algorithm, by ensuring all features are weighted similarly. The standardization method was applied to each of the 100 feature sets, so that they each have zero mean and unit standard deviation [25,26]. The normalized features are calculated as $z = (x - \mu)/\sigma$, where x are the unnormalized features, and μ and σ are the mean and standard deviation of the corresponding feature set, respectively. Note that μ and σ are calculated from the training instances only, but are used for normalizing both training and testing feature sets.

### Classification

We experimented with several classifiers [27] for classification of 1p/19q status:

- support vector machine (SVM) classifier [28,29];
- random forest (RF) classifier [30];
- k-nearest neighbors (KNN) classifier [31];
- extra-trees (ET) classifier [32,33];
- gradient boosting (GB) classifier [34];
- adaptive boosting (AB) classifier [35];
- stochastic gradient descent (SGD) classifier [36,37];
- histogram-based gradient boosting (HGB) classification tree classifier [38];
- extreme gradient boosting (XGB) classifier [39];
- logistic regression (LR) classifier [40];
- decision tree (DT) classifier [41,42];
- Gaussian naive Bayes (GNB) classifier [43]; and
- linear discriminant analysis (LDA) classifier [44].

A list of classes and input parameters used for implementing each classifier is shown in Table 2. Experiments with the above classifiers were performed using implementations from the scikit-learn 0.24.2 (sklearn) library [26], except for the XGB classifier, which was implemented using the XGBoost 1.7.1 (xgboost) library [39].

We also experimented with bagging [45], an ensemble meta-estimator in which random subsets of the training dataset are used to fit multiple instances of a base estimator, and then the final prediction is obtained by aggregating the individual predictions from each estimator instance (Figure 3). The default base estimator in the scikit-learn implementation of bagging is a decision tree classifier, but the user may specify a different classifier as base estimator. We experimented with bagging using each of the other methods listed in Table

2 as the base estimator. We used the BaggingClassifier class from the sklearn library with default input parameters, except for n_estimators=200 and base_estimator specified as one of the classifiers from Table 2.

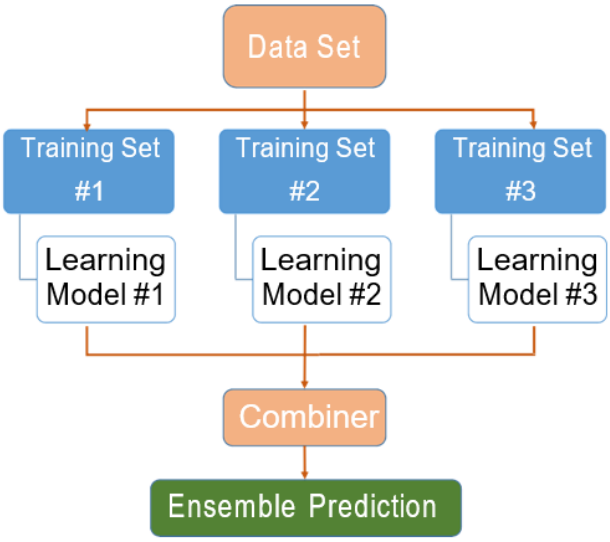


**Figure 3.** Simplified process diagram of parallel ensemble classifying methods such as bagging.

**Table 2.** List of classes and input parameters used for implementing each classifier.

| Classifier | Class and input parameters |
|---|---|
| SVM | SVC(kernel='linear', probability=True) |
| RF | RandomForestClassifier(n_estimators=200) |
| KNN | KNeighborsClassifier(n_neighbors=3,weights='uniform') |
| ET | ExtraTreesClassifier(n_estimators=200) |
| GB | GradientBoostingClassifier(n_estimators=200,max_depth=1) |
| AB | AdaBoostClassifier(n_estimators=200) |
| SGD | CalibratedClassifierCV(base_estimator=SGDClassifier(penalty='elasticnet', loss='log',learning_rate='optimal',eta0=1,alpha= 0.1)) |
| HGB | HistGradientBoostingClassifier(max_bins=255,max_iter=100) |
| XGB | XGBClassifier(n_estimators=200,eval_metric='logloss') |
| LR | LogisticRegression(class_weight='balanced', solver='newton-cg') |
| DT | DecisionTreeClassifier(max_features=30) |
| GNB | GaussianNB() |
| LDA | LinearDiscriminantAnalysis() |

Note: the XGBClassifier class is from the xgboost library; all other classes are from the sklearn library; default input parameters were used where not specified above.

## Evaluating the performance of the classifiers

The following widely-accepted metrics [46] were used to evaluate the performance of each classifier in predicting 1p/19q status: balanced classification accuracy (BCA), accuracy (Acc), sensitivity or recall (Se), specificity (Spe), precision (Pre), F1 score (F1), and area under curve (AUC).

Each classifier was simulated 100 times, with each new simulation beginning with a new train-test data split, followed by feature normalization. For each classifier, the metrics listed above were assessed for each simulation, and then averaged. Confidence intervals for each metric and for each classifier were also calculated — at the 95% confidence level, based on a two-tailed Student's t-distribution — from the results from these 100 simulations.

The AUC values represent the area under the receiver operating characteristic (ROC) curve [47], and were obtained while executing the classification step, as they were calculated using the roc_auc_score implementation from the sklearn library.

The AUC measures the probability that a positive sample will be labeled as positive by the classifier.

The other metrics were calculated after the classification step, from the confusion matrices obtained from each simulation, as:

$$BCA = ((Se + SPE) / 2) \quad (1)$$

$$Acc = ((TP + TN) / (TP + FP + TN + FN)) \quad (2)$$

$$Se = (TP / (TP + FN)) \quad (3)$$

$$Spe = (TN / (TN + FP)) \quad (4)$$

$$Pre = (TP / (TP + FP)) \quad (5)$$

$$F1 = 2 \cdot ((Pre \cdot Se) / (Pre + Se)) \quad (6)$$

where TP, FP, TN, and FN represent the number of true positives, false positives, true negatives, and false negatives, respectively, obtained from each simulation. Codeleted instances were treated as “positives”, and non-deleted instances were treated as “negatives”. All metrics were converted to percentage values, by multiplying each by 100%.

Note that the BCA is the arithmetic mean of sensitivity and specificity. This metric is particularly useful when dealing with imbalanced data [48], as is our case. Also note that the F1 score is the harmonic mean of precision and sensitivity. As such, a very low value of precision yields a very low F1 score, even if the sensitivity is very high; similarly, a very low value of sensitivity yields a very low F1 score, even if the precision is very high.

### Comparing different methods

Two different tumors may have similar size, shape and/or pixel gray-level distribution (these would be quantified by shape-based and first order statistics features), but may differ in spatial interrelationship of pixel intensities (which would be quantified by texture-based features). Texture analysis may be especially useful when dealing with markedly heterogeneous tumors [24]. Quantitative texture features extracted from T2 images have been shown to predict 1p/19q codeletion with high sensitivity and specificity [49].

Thus, we experimented with sets of 68 and 100 radiomic features. First order statistics and shape features were not used when working with only 68 features, i.e., only texture-based features were used (see Table 1). The average of each metric for each classifier, obtained with 68 features, was compared with that obtained with 100 features.

We also experimented with bagging using each of the other methods listed in Table 2 as the base estimator, and then compared each metric for each classifier with that obtained without bagging. This was done because default (DT-based) bagging showed very promising results in our initial experiments, so we decided to investigate whether this ensemble approach could improve the performance of other classifiers.

In each of those comparisons, we performed a two-sample t-test using GNU Octave 7.2.0. When $p < 0.05$, the null hypothesis of equal means was rejected. Normality tests were not performed, as our sample size ($n = 100$) is large [50].

## RESULTS

Table 3 presents the performance results obtained with each of the classifiers listed in Table 2, using 68 texture-based radiomic features (features associated with first order statistics and shape were not used in this experiment; see Table 1). Each entry corresponds to a 100-simulation average, and the best result for each metric is highlighted in bold. In this first experiment, SVM and RF presented the best overall results, and displayed similar performances. SVM presented the highest BCA (77.3%), while RF presented the highest accuracy (80.3%) and AUC (84.6%). These classifiers also presented competitive sensitivity, specificity, precision and F1 score. The SGD classifier presented higher sensitivity (95.7%) than SVM and RF, but provided very low specificity (51.5%). This is because SGD

presented few false negatives (codeleted instances classified as non-deleted), but many false positives (non-deleted instances classified as codeleted). The LR classifier provided the highest specificity (66.5%) and the lowest sensitivity (75.7%), as it presented fewer false positives, but more false negatives.

**Table 3**. Classification performance using 68 texture-based radiomic features. Each entry corresponds to a 100-simulation average. The best result for each metric is highlighted in bold.

| Classifier | BCA | Acc | Se | Spe | Pre | AUC | F1 |
|---|---|---|---|---|---|---|---|
| SVM | **77.3** | 79.8 | 91.0 | 63.5 | 79.5 | 80.0 | 84.2 |
| RF | 76.0 | **80.3** | 90.9 | 61.1 | 81.2 | **84.6** | 85.2 |
| KNN | 75.3 | 79.3 | 91.4 | 59.1 | 78.7 | 77.8 | 84.0 |
| ET | 74.8 | 79.1 | 90.1 | 59.6 | 79.5 | 84.2 | 83.9 |
| GB | 74.8 | 79.2 | 90.8 | 58.8 | 79.9 | 82.9 | 84.4 |
| AB | 74.5 | 79.4 | 88.4 | 60.5 | **81.4** | 82.0 | 84.2 |
| SGD | 73.6 | **80.3** | **95.7** | 51.5 | 78.9 | 79.3 | **86.0** |
| HGB | 72.9 | 76.4 | 84.6 | 61.1 | 79.8 | 79.9 | 81.4 |
| XGB | 72.5 | 76.0 | 85.9 | 59.0 | 78.1 | 77.7 | 81.1 |
| LR | 71.1 | 71.6 | 75.7 | **66.5** | 78.3 | 78.8 | 76.1 |
| DT | 68.8 | 71.4 | 77.5 | 60.0 | 78.8 | 68.8 | 77.2 |
| GNB | 67.9 | 72.3 | 85.8 | 49.9 | 74.5 | 70.9 | 79.0 |
| LDA | 61.9 | 66.2 | 77.3 | 46.5 | 72.6 | 64.3 | 73.8 |

Table 4 presents the results obtained with the same classifiers, but using 100 radiomic features (i.e., features associated with first order statistics and shape were not excluded). Statistically-significant differences were found for some metrics and some classifiers. The use of additional radiomic features resulted in worse performance (significantly lower average for at least one metric, as indicated with $*$) for most classifiers: SVM, RF, KNN, ET, GB, AB, SGD, and LDA. We did not observe any significant difference in performance for these classifiers: HGB, XGB, and DT. We observed improved results (indicated with $**$) for only two classifiers: LR, for which the F1 score was significantly improved from 76.1% to 79.5%, and GNB, for which the AUC was significantly improved from 70.9% to 75.7%. Note GNB still presents very poor overall performance compared to the other classifiers from Tables 3 and 4, while LR still presents high specificity, but low sensitivity. Based on these results, we used the 68-feature set in our next experiment, instead of the full 100-feature set.

Table 5 presents the results obtained with the same classifiers, but this time using each classifier as the base estimator of the bagging classifier. The 68-feature set was used. The use of bagging significantly improved all metrics (indicated with $**$) for the XGB and DT classifiers, while significantly worsened all metrics (indicated with $*$) for the KNN classifier. Bagging also significantly improved the sensitivity of SVM (from 91.0% to 96.3%), but at the cost of significantly lower specificity (from 63.5% to 39.1%). HGB presented an important improvement in sensitivity (from 84.6% to 90.2%), which resulted in significant improvement in BCA, accuracy, and F1 score. We did not observe any significant difference in performance for these classifiers: RF, ET, GB, and GNB. Most notably, bagging with DT presented the highest BCA (79.3%), accuracy (81.9%), and specificity (67.8%) of all our experiments; our second highest precision (82.6%), AUC (84.2%), and F1 score (85.9%); and competitive sensitivity (90.9%). We consider DT-based bagging to be our best overall classifier. Note that bagging with XGB and bagging with HGB also presented excellent results.

Table 6 summarizes the results we obtained with our best overall classifier: bagging method with a decision tree base estimator, using a 68-feature set. Mean, standard deviation and 95% confidence intervals are shown for confusion matrix values and for all classification metrics. The confusion matrix values show that the average number of misclassifications (FN and FP) is small, compared to the number of correct classifications (TP and TN). Particularly, the number of false negatives is very small compared to the number of true positives. This ensures high sensitivity (90.9%). Unfortunately, the number of false positives is not as low when compared with the number of true negatives. This hurts specificity. Still, our best specificity result (67.8%) was obtained using DT-based bagging.

**Table 4.** Classification performance using the full 100-feature set. Each entry corresponds to a 100-simulation average. The best result for eachmetric is highlighted in bold.

| Classifier | BCA | Acc | Se | Spe | Pre | AUC | F1 |
|---|---|---|---|---|---|---|---|
| SVM | 71.8* | 75.6* | 85.3* | 58.2 | 79.2 | 73.5* | 81.3* |
| RF | 74.6 | 78,1 | 88.8 | 60.5 | 79.9 | 81.0* | 83.4 |
| KNN | 64.2* | 69.4* | 82.9* | 45.6* | 74.3* | 70.3* | 77.2* |
| ET | 73.5 | 77.7 | 89.9 | 57.1 | 78.4 | 80.3* | 83.1 |
| GB | **74.7** | 77.6 | 85.7* | 63.6 | 80.3 | 80.9 | 82.2 |
| AB | 70.4* | 74.1* | 82.7* | 58.1 | 78.5 | 74.7* | 79.6* |
| SGD | 72.8 | **78.5** | **93.3*** | 52.3 | 77.7 | 79.3 | **84.2** |
| HGB | 71.4 | 75.8 | 85.8 | 57.1 | 79.3 | 77.5 | 81.7 |
| XGB | 73.2 | 76.7 | 86.8 | 59.6 | 79.8 | **81.1** | 82.4 |
| LR | 72.4 | 74.6 | 78.8 | **66.0** | **81.3** | 77.9 | 79.5** |
| DT | 67.6 | 69.9 | 76.0 | 59.2 | 78.1 | 67.3 | 76.0 |
| GNB | 69.0 | 73.6 | 84.9 | 53.0 | 76.4 | 75.7** | 79.7 |
| LDA | 61.8 | 64.5 | 71.2* | 52.4 | 73.0 | 61.5 | 71.0 |

∗ $p < 0.05$ and mean is lower than with 68 features.
∗∗ $p < 0.05$ and mean is higher than with 68 features.

**Table 5.** Classification performance using each classifier as the base estimator of thebagging classifier. The 68-feature set was used. Each entry corresponds to a 100-simulation average. The best result for each metric is highlighted in bold.

| Base Estimator | BCA | Acc | Se | Spe | Pre | AUC | F1 |
|---|---|---|---|---|---|---|---|
| SVM | 67.7* | 76.1* | **96.3**** | 39.1* | 74.4* | 78.0 | 83.4 |
| RF | 75.9 | 79.2 | 90.1 | 61.8 | 79.8 | 81.7 | 83.9 |
| KNN | 61.4* | 56.1* | 76.0* | 46.7* | 71.9* | 68.6* | 73.2* |
| ET | 72.4 | 78.3 | 89.6 | 55.2 | 80.8 | 82.5 | 84.4 |
| GB | 76.9 | 81.2 | 81.0 | 62.8 | 81.8 | 83.3 | 85.6 |
| AB | 76.8 | 80.3 | 92.5** | 61.1 | 79.8 | 82.9 | 85.0 |
| SGD | 76.1 | 79.4 | 87.8* | 64.3** | 81.5 | 80.7 | 83.8 |
| HGB | 77.9** | 80.9** | 90.2** | 65.6 | 81.9 | 83.1 | 85.3** |
| XGB | 78.5** | 81.6** | 90.3** | 66.7** | **83.0**** | 82.8** | 85.8** |
| LR | 69.5 | 71.9 | 79.9** | 59.2* | 76.7 | 76.2 | 77.4 |
| DT | **79.3**** | **81.9**** | 90.9** | **67.8**** | 82.6** | **84.2**** | **85.9**** |
| GNB | 68.4 | 73.8 | 84.0 | 52.8 | 77.4 | 72.5 | 79.9 |
| LDA | 61.7 | 65.6 | 78.3 | 45.2 | 71.2 | 71.5** | 73.4 |

∗ $p < 0.05$ and mean is lower than without bagging.
∗∗ $p < 0.05$ and mean is higher than without bagging.

**Table 6.** Summary of results obtained with our best overall classifier: bagging method with a decision tree base estimator, using a 68-feature set. Mean, standard deviation and 95% confidence intervals are shown for confusion matrix values and for all classification metrics

| | mean ± std dev | confidence interval |
|---|---|---|
| TP | 9.32 ± 1.86 | 8.96 – 9.68 |
| FN | 0.94 ± 0.93 | 0.76 – 1.12 |
| TN | 3.78 ± 1.35 | 3.51 – 4.05 |
| FP | 1.96 ± 1.36 | 1.69 – 2.23 |
| BCA | 79.3 ± 11.3 | 77.1 – 81.5 |
| Acc | 81.9 ± 9.9 | 79.9 – 83.8 |
| Se | 90.9 ± 8.9 | 89.2 – 92.7 |
| Spe | 67.8 ± 20.0 | 63.8 – 71.7 |
| Pre | 82.6 ± 12.0 | 80.2 – 84.9 |
| AUC | 84.2 ± 10.0 | 82.2 – 86.2 |
| F1 | 85.9 ± 8.3 | 84.3 – 87.6 |

Figure 4 presents the ROC curve obtained by averaging over 100 simulations of our DT-based bagging classifier. Metrics such as sensitivity, specificity, and precision are dependable on the classifier's discrimination threshold, so it is often useful to analyze the ROC curve, which illustrates how the classification results change as the classifier's discrimination threshold is varied. The ROC curve is a plot of true positive

rate (Se) against false positive rate (1 − Spe). The ROC curve is computed based on probability estimates of the positive class (codeleted 1p/19q status). The area under the average ROC curve was 0.842, which corresponds to the mean AUC value of 84.2%. This is an estimate of the probability of accurately labeling a positive (codeleted 1p/19q status) sample.

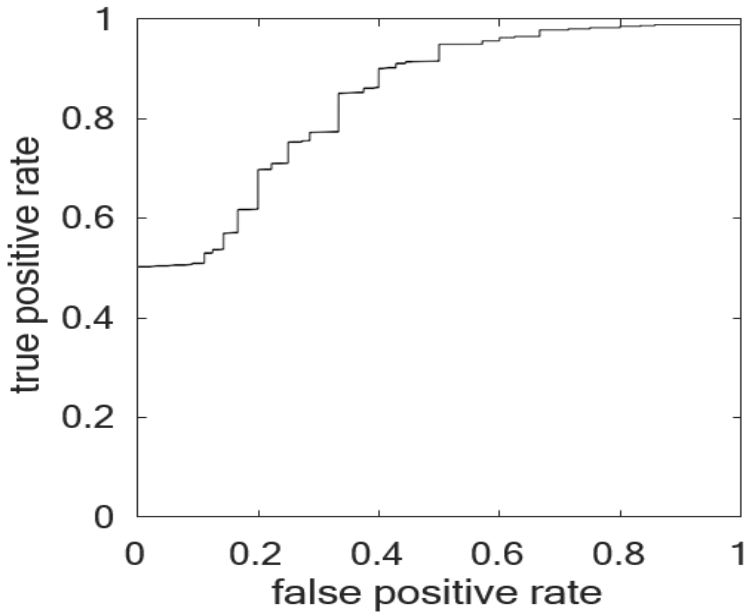


**Figure 4**. ROC curve obtained by averaging over 100 simulations of our DT-based bagging classifier. The area under the ROC curve was 0.84.

## DISCUSSION

Our experiments evaluated the potential of shape-based, first order statistics, and texture-based radiomic features for machine learning classification of 1p/19q chromosomal codeletion status in low-grade gliomas. Shape-based features quantify the tumor's shape and size; first order statistics features quantify the tumor's pixel intensity distribution (or histogram); and texture-based features (associated with second and higher order statistics) quantify pixel interrelationships, such as spatial relationships and co-occurrence of the pixel values [24]. The PyRadiomics library proved adequate for extracting radiomic features from the LGG-1p19qDeletion dataset, requiring only a file format conversion (from NIfTI to NRRD). We demonstrated promising classification results using exclusively texture-based radiomic features extracted from T2-weighted MRI images. Texture potentially correlate with cellular density, angiogenesis, and necrosis; hence, texture analysis can be used to measure tumor heterogeneity and to predict tumor biology [24]. For example, oligodendrogliomas usually present calcifications and bleeding, while astrocytomas tend to be more homogeneous. Thus, heterogeneity is an important visual aspect that specialist evaluate on MRI images when attempting to differentiate between these types of tumors. Texture analysis has been shown to predict 1p/19q codeletion with high sensitivity and specificity [49]. Our experiments showed that not only texture-based features are sufficient for classification of 1p/19q status, but including radiomic features associated with first order statistics and shape does not improve the performance of the classifiers, and in many cases worsens it.

Ensemble methods such as bagging and boosting have been shown to provide better accuracy than single base classifiers in several applications [51-54]. Single base estimators often have a high bias or too much variance. Ensemble methods try to reduce bias and/or variance by combining several learners [55-57]. In bagging, several homogeneous learners are trained independently, in parallel (as illustrated in Figure 3); in boosting, the homogeneous learners are trained sequentially. We experimented with several boosting-based classifiers (GB, AB, HGB, and XGB), but our best results were obtained using bagging-based classifiers. Interestingly, the use of bagging improved the performance of some of the boosting classifiers (see Table 5). Our best result, however, was obtained with a bagging classifier using the decision tree algorithm as base estimator.

Table 7 presents a comparison between the proposed DT-based bagging classifier and several methods from the literature. These represent the state-of-the-art in 1p/19q codeletion status classification based on radiomics and machine learning. Table 7 also includes metrics from a study in which two neurosurgeons and two neuroradiologists evaluated the T2-weighted images from the LGG-1p19qDeletion dataset [13]. The evaluators rated each image with respect to whether they were able to confidently determine the tumor's 1p/19q status, using a scale of 1 to 5 (where 1 corresponded to "unsure", and 5 corresponded to "confident"). This score was converted into a codeletion prediction marker, from which the presented metrics were calculated. The comparison in Table 7 shows that our DT-based bagging classifier provides accuracy and AUC values comparable to the best methods from the literature. If we focus on works that used exclusively the same database (Refs. [14] and [16]), our classifier shows better sensitivity, but lower specificity. Our method also achieved better sensitivity and AUC than that provided by neurosurgeons and neuroradiologists analyzing the same images.

**Table 7**. Comparison with state-of-the-art methods from the literature for 1p/19q status classification based on radiomics and machine learning.

| Reference | Dataset | Classifier | BCA | Acc | Se | Spe | Pre | AUC | F1 |
|---|---|---|---|---|---|---|---|---|---|
| [13] | LGG-1p19qDeletion | neurosurgeons* | | 48.9 | 41.5 | 63.8 | | 51.5 | |
| | | neuroradiologists* | | 63.2 | 53.5 | 81.8 | | 81.1 | |
| | Own - EMC/HMC (train)<br>LGG-1p19qDeletion (test) | SVM | | 69.3 | 73.2 | 61.7 | 78.7 | 72.3 | 69.7 |
| [14] | LGG-1p19qDeletion | quadratic SVM | | 80.0 | 66.7 | **100.0** | | | |
| [15] | Own | RF | | 70.0 | 68.3 | 71.2 | | 76.0 | |
| [16] | LGG-1p19qDeletion | multilayer perceptron | | **83.8** | 87.5 | 75.8 | **90.5** | 86.9 | **88.3** |
| [17] | Own – PUMCH | RF | | 83.3 | **100.0** | 75.0 | | **88.9** | |
| [18] | TCGA-LGG | XGB | | | 75.0 | 85.0 | | 80.0 | |
| [19] | LGG-1p19qDeletion<br>TCGA-LGG BraTS 2019 | CNN | | 83.3 | 86.7 | 79.2 | | 82.1 | |
| | Own - GUH | | | 75.0 | 58.3 | 82.1 | | 86.6 | |
| proposed | LGG-1p19qDeletion | DT-based bagging | 79.3 | 81.9 | 90.9 | 67.8 | 82.6 | 84.2 | 85.9 |

* Ref. [13] presents a comparative study in which two neurosurgeons and two neuroradiologists rated the dataset's images with respect to whether they were able to confidently determine the tumor's 1p/19q status, using a scale of 1 to 5 (where 1 corresponded to "unsure", and 5 corresponded to "confident"). This score was converted into a codeletion prediction marker, from which these metrics were calculated.

Akkus and coauthors [20] reported 87.7% accuracy using a CNN-based method applied onto the LGG-1p19qDeletion dataset. We decided not to include this result in Table 7 because it was obtained by using three slices from each subject, and then dividing the slices into training and testing sets, such that slices from the same subject may have been used for both training and testing steps. We feel that this may configure data leakage.

When assessing 1p/19q status, it is more critical to avoid false negatives (misdiagnosing a codeleted tumor) than it is to avoid false positive (misdiagnosing a non-deleted tumor); i.e., high sensitivity is more critical than high specificity. Positive codeletion diagnosis is important because patients with codeleted tumors have better prognosis and longer survival, as these tumors respond well to dedicated chemotherapy (based on procarbanize, lomustine and vincristine), and to radiotherapy [4]. Our DT-based bagging classifier achieved higher sensitivity (90.9%) than most methods from the literature. Some of our classifiers achieved even higher sensitivity (as high as 96.3%), at the cost of lower specificity. While none of the classifiers we evaluated achieved specificity higher than 67.8%, we believe this may be due to the small size of the LGG-1p19qDeletion dataset, particularly the small number of non-deleted instances. Using an imbalanced dataset is often an issue, as the data patterns of the majority class may stand out from those of the minority class. A model trained with an imbalanced dataset may project higher accuracy onto the majority class, and possibly display poor classification performance for the minority class [58]. Aiming to address this issue, we experimented with minority oversampling techniques such as the adaptive synthetic algorithm [59] and the synthetic minority oversampling technique [60], but this did not improve the classification results. We also tried to match the number of instances of each class in the training set by reducing the number of instances from the majority class [61], but this also proved unfruitful. Experiments with dimensionality reduction techniques [62] also showed no improvement in classification performance. Li and coauthors [63] suggest a relation of 10 to 15 instances per radiomic feature used in the classification stage. Thus, we strongly feel that our classifiers could have achieved better performance if a larger dataset was available.

## CONCLUSION

This work evaluated several machine learning classifiers for the non-invasive evaluation of the 1p/19q chromosome codeletion status in low grade gliomas, using radiomic features extracted from T2-weighted magnetic resonance images. Our experiments showed that including radiomic features associated with first order statistics and shape does not improve the performance of the classifiers, and in many cases worsens it. Our best results were achieved using a bagging ensemble estimator based on the decision tree classifier, using only texture-based radiomic features. Compared to other works that used the same database, this approach provided higher sensitivity. It also achieved higher sensitivity and AUC than that provided by neurosurgeons and neuroradiologists analyzing the same images. Results could be improved by using a larger dataset for training the classifiers.

Patients with 1p/19q codeletion have better survival rates and prognosis than non- deleted patients, but the molecular assessment of brain tumors through biopsies is an invasive procedure, and is subject to sampling errors. Therefore, the techniques presented in this work have strong potential for aiding in better clinical, surgical, and therapeutic decision-making.

**Funding:** This research received no external funding.
**Acknowledgments:** The authors would like to express thanks to Federal Institute Goiano.
**Conflicts of Interest:** The authors declare no conflict of interest.

## REFERENCES

1. Dubbink HJ, Atmodimedjo PN, Kros JM, French PJ, Sanson M, Idbaih A, et al. Molecular classification of anaplastic oligodendroglioma using next-generation sequencing: a report of the prospective randomized eortc brain tumor group 26951 phase iii trial. Neuro-oncology. 2015 jun;18(3):388–400. doi.org/10.1093/neuonc/nov182.
2. Louis DN, Perry A, Reifenberger G, Von Deimling A, Figarella-Branger D, Cavenee WK, et al. The 2016 world health organization classification of tumors of the central nervous system: a summary. Acta Neuropathol. 2016 mai;131(6):803–20. doi.org/10.1007/s00401-016-1545-1.
3. Weller M, Van Den Bent M, Tonn JC, Stupp R, Preusser M, Cohen-Jonathan-Moyal E, et al. European association for neuro-oncology (EANO) guideline onthe diagnosis and treatment of adult astrocytic and oligodendroglial gliomas. Lancet Oncol. 2017 jun;18(6):315–29. doi.org/10.1016/S1470-2045(17)30194-8.
4. Eckel-Passow JE, Lachance DH, Molinaro AM, Walsh KM, Decker PA, Sicotte H, et al. Glioma groups based on 1p/19q, IDH, and TERT promoter mutations in tumors. New Eng. J. Med. 2015 jun;372(26):2499–508. doi.org/10.1056/NEJMoa1407279.
5. Wijnenga MM, Mattni T, French PJ, Rutten GJ, Leenstra S, Kloet F, et al. Does early resection of presumed low-grade glioma improve survival? a clinical perspective. J. Neuro-oncol. 2017 apr;133(1):137–46. doi.org/10.1007/s11060-017-2418-8.
6. Cha S. Update on brain tumor imaging: from anatomy to physiology. Am. J. Neuroradiol. 2006 mar;27(3):475–87.
7. DeAngelis LM. Brain tumors. New Eng. J. Med. 2001 jan;344(2):114–23. doi.org/10.1056/NEJM200101113440207.
8. Lambin P, Leijenaar RT, Deist TM, Peerlings J, De Jong EE, Van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat. Rev. Clin. Oncol. 2017 oct;14(12):749–762. doi.org/10.1038/nrclinonc. 2017.141.
9. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology. 2015 nov;278(2):563-77. doi.org/10.1148/radiol.2015151169.
10. Ba stanlar Y, O¨ zuysal M. Introduction to machine learning in miRNomics: MicroRNA biology and computational analysis. Humana Press. 2014 mai;105–28.
11. Alcaide-Leon P, Dufort P, Geraldo A, Alshafai L, Maralani P, Spears J, et al. Differentiation of enhancing glioma and primary central nervous system lymphoma by texture-based machine learning. Am. J. Neuroradiol. 2017 jun;38(6):1145–50. doi.org/10.3174/ajnr.A5173.
12. Zhang Y, Chen C, Cheng Y, Teng Y, Guo W, Xu H, et al. Ability of radiomics in differentiation of anaplastic oligodendroglioma from atypical low-grade oligodendroglioma using machine-learning approach. Frontiers in Oncol. 2019 dec;9:1371. doi.org/10.3389/fonc.2019.01371.
13. van der Voort SR, Incekara F, Wijnenga MM, Kapas G, Gardeniers M, Schouten JW, et al. Predicting the 1p/19q codeletion status of presumedlow-grade glioma with an externally validated machine learning algorithm. Clin. Cancer Res. 2019 dec; 25(24):7455–62. doi.org/10.1158/1078-0432.CCR-19-1127.
14. Lu CF, Hsu FT, Hsieh KLC, Kao YCJ, Cheng SJ, Hsu JBK, et al. Machine learning - based radiomics for molecular subtyping of gliomas. Clin. Cancer Res. 2018 sep;24(18):4429–36. doi.org/10.1158/1078-0432.CCR-17-3445.

15. Han Y, Xie Z, Zang Y, Zhang S, Gu D, Zhou M, et al. Non-invasive genotype prediction of chromosome 1p/19q co-deletion by development and validation of an MRI-based radiomics signature in lower-grade gliomas. J. Neuro-oncol. 2018 aug;140(2):297–306.doi.org/10.1007/s11060-018-2953-y.
16. Kocak B, Durmaz ES, Ates E, Sel I, Turgut GS, Kaya OK, et al. Radiogenomics of lower-grade gliomas: Machine learning-based MRI texture analysis for predicting 1p/19q codeletion status. Eur. Radiol. 2019 nov;30(2):877–86. doi.org/10.1007/s00330-019-06492-2.
17. Kong Z, Jiang C, Zhang Y, Liu S, Liu D, Liu Z, et al. Thin-slice magnetic resonance imaging-based radiomics signature predicts chromosomal 1p/19q co-deletion status in grade II and III gliomas. Frontiers in Neurol. 2020 oct;11:551771. doi.org/10.3389/fneur.2020.551771.
18. Shboul ZA, Chen J, M Iftekharuddin K. Prediction of molecular mutations in diffuse low-grade gliomas using MR imaging features. Sci. Rep-UK. 2020 feb;10(1):1–13. doi.org/10.1038/s41598-020-60550-0.
19. Decuyper M, Bonte S, Deblaere K, Van Holen R. Automated MRI based pipeline for segmentation and prediction of grade, IDH mutation and 1p19q co-deletion in glioma. Comput. Med. Imag. Grap. 2021 mar; 88:101831. doi.org/10.1016/j.compmedimag.2020.101831.
20. Akkus Z, Ali I, Sedlář J, Agrawal JP, Parney IF, Giannini C, et al. Predicting deletion of chromosomal arms 1p/19q in low-grade gliomas from MR images using machine intelligence. J. Digit. Imaging. 2017 jun;30(4):469–476. doi.org/10.1007/s10278-017-9984-3.
21. Erickson B, Akkus Z, Sedlar J, Korfiatis P. Data from LGG-1p19q deletion. Cancer Imag. Arch. 2017. doi.org/10.7937/K9/TCIA.2017.dwehtz9v.
22. Van Griethuysen JJ, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. Cancer Res. 2017 nov;77(21):104–107. doi.org/10.1158/0008-5472.CAN-17-0339.
23. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, et al. 3D slicer as an image computing platform for the Quantitative Imaging Network. Magn. Reson. Imaging. 2012 nov;30(9):1323–41. doi.org/10.1016/j.mri.2012.05.001.
24. Soni N, Priya S, Bathla G. Texture analysis in cerebral gliomas: a review of the literature. Am. J. Neuroradiol. 2019 jun;40(6):928–34. doi.org/10.3174/ajnr.A6075.
25. Erwin K. Advanced Engineering Mathematics. 4th ed. London. John Wiley, Inc. 1979. p. 898.
26. Geron A. Hands-on Machine Learning with Scikit-learn and Tensorflow: Concepts Tools and Techniques to Build Intelligent Systems. 2nd ed. United States. O'Reilly Media.2019. p. 600.
27. Bonaccorso G. Machine Learning Algorithms: Popular Algorithms for Data Science and Machine Learning. 2nd ed. United States. Packt Publishing Ltd. 2018. P. 522.
28. Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. In: Proceedings of the Fifth Annual Workshop on Computational Learning Theory; 1992 jul 27-29; Pittsburgh Pennsylvania, USA. ACM; 1992. p. 144–152. doi.org/10.1145/130385.130401.
29. Cortes C, Vapnik V. Support-vector networks. Mach. Learn.1995 sep;20(3):273–97. doi.org/10.1007/BF00994018.
30. Breiman L. Random forests. Mach. Learn. 2001 oct;45(1):5–32. doi.org/10.1023/A:1010933404324.
31. Wu Y, Ianakiev K, Govindaraju V. Improved k-nearest neighbor classification. Pattern Recogn. 2012 oct;35(10):2311–2318. doi.org/10.1016/S0031-3203(01)00132-7.
32. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Mach. Lear. 2006 mar;63(1):3–42. doi.org/10.1007/s10994-006-6226-1.
33. Bhati BS, Rai C. Ensemble based approach for intrusion detection using extra tree classifier. In: Intelligent Computing in Engineering; 2020 apr; Singapure, Asia. Springer; 2020. p. 213–220. doi.org/10.1007/978-981-15-2780-7_25.
34. Friedman JH. Greedy function approximation: a gradient boosting machine [internet]. Annals of statistics; 2001; JSTOR; 2001. p. 1189–1232. Available from: https://www.jstor.org/stable/2699986.
35. Schapire RE. Explaining adaboost. In: Empirical Inference; 2013; Berlin, Heidelberg. Springer; 2013. p. 37–52. doi.org/10.1007/978-3-642-41136-6_5.
36. Amari SI. Backpropagation and stochastic gradient descent method. Neurocomputing. 1993 jun;5(4-5):185–96. doi.org/10.1016/0925-2312(93)90006-O.
37. Bottou L. Large-scale machine learning with stochastic gradient descent. In: Proceedings of COMPSTAT'2010; 2010 aug 22-27; Paris, France. Physica-Verlag HD; p. 177–86. doi.org/10.1007/978-3-7908-2604-3_16.
38. Shafer J, Agrawal R, Mehta M. SPRINT: A scalable parallel classifier for data mining [internet]. In: Proceedings of the 22nd VLDB Conference; 1996; Mumbai, India. 1996. p. 544–555. Available from: http://www.cs.ecu.edu/dingq/CSCI6905/readings/shafer96sprint.pdf.
39. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining; 2016 aug; [place unknown]. ACM; 2016. p. 785–794. doi.org/10.1145/2939672.2939785.
40. LaValley MP. Logistic regression. Circulation. 2008 may;117(18):2395–2399. doi.org/10.1161/CIRCULATIONAHA.106.682658.
41. Freund Y, Mason L. The alternating decision tree learning algorithm. In: Icml; 1999; [place unknown]. p.124–133. Available from https://staff.icar.cnr.it/manco/Teaching/2006/datamining/articoli/Freund_Atrees.pdf.

42. Su J, Zhang H. A fast decision tree learning algorithm. In: American Association for Artificial Intelligence; 2006; [place unknown]. p. 500–505. Available from: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=ec2fa7d5a29e91c50362f0a87372f027ece2b86d
43. Rish I, et al. An empirical study of the naive bayes classifier. In: IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence; 2001; [place unknown]. p. 41–46. Available from: https://www.cc.gatech.edu/home/isbell/classes/reading/papers/Rish.pdf.
44. Balakrishnama S, Ganapathiraju A. Linear discriminant analysis-a brief tutorial. Inst. Sig. Inf. Process. 1998. 18:1–8.
45. Breiman L. Bagging predictors. Mach. Learn. 1996 aug;24(2):123–40. doi.org/10.1007/BF00058655.
46. Dalianis H. Evaluation metrics and evaluation. In: Clinical Text Mining: secondary use of electronic patient records; 2018 may; [place unknown]. Springer; 2018. p. 45-53. doi.org/10.1007/978-3-319-78503-5_6.
47. Saxena A, Celaya J, Balaban E, Goebel K, Saha B, Saha S, et al. Metrics for evaluating performance of prognostic techniques. In: 2008 International Conference on Prognostics and Health Management; 2008 dec 8-9; Denver, CO, USA. IEEE; 2008. p. 1–17. doi.org/10.1109/PHM.2008.4711436.
48. Brodersen KH, Ong CS, Stephan KE, Buhmann JM. The balanced accuracy and its posterior distribution. In: 2010 20th International Conference on Pattern Recognition; 2010 aug 23-26; Istanbul, Turkey. IEEE; 2010. p. 3121–4. doi.org/10.1109/ICPR.2010.764.
49. Brown R, Zlatescu M, Sijben A, Roldan G, Easaw J, Forsyth P, et al. The use of magnetic resonance imaging to noninvasively detect genetic signatures in oligodendroglioma. Clin. Cancer Res. 2008 apr;14(8):2357–62. doi.org/10.1158/1078-0432.CCR-07-1964.
50. Fagerland MW. t-tests, non-parametric tests, and large studies—a paradox of statistical practice? BMC Med. Res. Methodol. 2012 jun;12(1):1–7. doi.org/10.1186/1471-2288-12-78.
51. El Aissaoui O, El Madani Y, Oughdir L, El Allioui Y. Combining supervised and unsupervised machine learning algorithms to predict the learners' learning styles. Procedia Comput. Sci. 2019 jul;148:87–96. doi.org/10.1016/j.procs.2019.01.012.
52. Viswanath SE, Chirra PV, Yim MC, Rofsky NM, Purysko AS, Rosen MA, et al. Comparing radiomic classifiers and classifier ensembles for detection of peripheral zone prostate tumors on t2-weighted mri: A multi-site study. BMC Med. Imaging. 2019 feb;19(1):1–12. doi.org/10.1186/s12880-019-0308-6.
53. Lavanya D, Rani KU. Ensemble decision tree classifier for breast cancer data. Int. J. Inf. Technol. Conv. Serv. 2012;2(1):17–24.
54. KP MN, Thiyagarajan P. Alzheimer's classification using dynamic ensemble of classifiers selection algorithms: A performance analysis. Biomed. Signal Proces. 2021 jul;68:102729. doi.org/10.1016/j.bspc.2021.102729.
55. Idri A, Hosni M, Abran A. Systematic mapping study of ensemble effort estimation. In Proceedings of the11th International Conference on Evaluation of Novel Software Approaches to Software Engineering (ENASE2016); 2012; ECITPRESS; 2012. p.132–139. Available from: https://pdfs.semanticscholar.org/c152/993c9e146fd44ac06cc4e92ed1cf59ca8f03.pdf.
56. Zhou ZH. Ensemble Methods: Foundations and Algorithms. 1st ed. London. CRC press, 2012. p. 236.
57. Dietterich TG. Ensemble methods in machine learning. In: International Workshop on Multiple Classifier Systems (MSC 2000); 2000 jun 21-23; Cagliari, Italy. Springer Berlin Heidelberg; 2000. p. 1–15. doi.org/10.1007/3-540-45014-9_1.
58. Ma Y, He H. Imbalanced learning: foundations, algorithms, and applications. 1st ed. New Jersey. Jonh Wiley and Sons, Inc. 2013. p. 216.
59. He H, Bai Y, Garcia EA, Li S. Adasyn: Adaptive synthetic sampling approach for imbalanced learning. In: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence); 2008 jun 01-08; Hong Kong, China. IEEE; 2008. p. 1322–1328. doi.org/10.1109/IJCNN.2008.4633969.
60. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. Smote: synthetic minority over-sampling technique. J. Artif. Intell. Res. 2002 jun;16:321–357. doi.org/10.1613/jair.953.
61. Zhang J, Chen L, Abid F. Prediction of breast cancer from imbalance respect using cluster-based undersampling method. J. Health. Eng. 2019 oct;2019:10. doi.org/10.1155/2019/7294582.
62. Ray S, Analytics B. Beginners guide to learn dimension reduction techniques [internet]. Analytics Vidhya; 2015 jul. Available from https://www.analyticsvidhya.com/blog/2015/07/dimension-reduction-methods/.
63. Li R, Xing L, Napel S, Rubin DL. Radiomics and Radiogenomics: Technical Basis and Clinical Applications. 1st ed. London. CRC press, 2021. p. 484.