European Journal of Radiology 139 (2021) 109678

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Research article

Development and external validation of a non-invasive molecular status predictor of chromosome 1p/19q co-deletion based on MRI radiomics analysis of Low Grade Glioma patients

Roberto Casale [a,*], Elizaveta Lavrova [a], Sebastian Sanduleanu [a], Henry C. Woodruff [a,b,1], Philippe Lambin [a,b,1]

[a] The D-Lab, Department of Precision Medicine, GROW- School for Oncology, Maastricht University Maastricht, the Netherlands
[b] Department of Radiology and Nuclear Medicine, GROW - School for Oncology, Maastricht University Medical Centre+, Maastricht, the Netherlands

ARTICLE INFO

Keywords:
Radiomics
MRI
Low grade glioma
1p/19q co-deletion
Cubic interpolation
Linear interpolation

ABSTRACT

Purpose: The 1p/19q co-deletion status has been demonstrated to be a prognostic biomarker in lower grade glioma (LGG). The objective of this study was to build a magnetic resonance (MRI)-derived radiomics model to predict the 1p/19q co-deletion status.

Method: 209 pathology-confirmed LGG patients from 2 different datasets from The Cancer Imaging Archive were retrospectively reviewed; one dataset with 159 patients as the training and discovery dataset and the other one with 50 patients as validation dataset.

Radiomics features were extracted from T2- and T1-weighted post-contrast MRI resampled data using linear and cubic interpolation methods.

For each of the voxel resampling methods a three-step approach was used for feature selection and a random forest (RF) classifier was trained on the training dataset. Model performance was evaluated on training and validation datasets and clinical utility indexes (CUIs) were computed. The distributions and intercorrelation for selected features were analyzed.

Results: Seven radiomics features were selected from the cubic interpolated features and five from the linear interpolated features on the training dataset. The RF classifier showed similar performance for cubic and linear interpolation methods in the training dataset with accuracies of 0.81 (0.75−0.86) and 0.76 (0.71−0.82) respectively; in the validation dataset the accuracy dropped to 0.72 (0.6−0.82) using cubic interpolation and 0.72 (0.6−0.84) using linear resampling. CUIs showed the model achieved satisfactory negative values (0.605 using cubic interpolation and 0.569 for linear interpolation).

Conclusions: MRI has the potential for predicting the 1p/19q status in LGGs. Both cubic and linear interpolation methods showed similar performance in external validation.

1. Introduction

Gliomas are tumors of the central nervous system and are the most frequent primary tumors arising in the brain [1]. They are classified into four grades based on their aggressiveness by The World Health Organization (WHO). WHO grade II (low grade) and grade III (anaplastic) diffuse gliomas form a heterogeneous group of neoplasms, also known as Low Grade Gliomas (LGGs), characterized by a wide range of malignant potential affecting mostly young adults [2]; LGG is potentially a fatal disease, with an median overall survival of around 7

Abbreviations: WHO, World Health Organization; LGG, low grade glioma; MRI, magnetic resonance imaging; SVM, support vector machine; RF, random forest; RFE, recursive feature elimination algorithm; CNN, convolutional neural networks; FISH, fluorescence in situ hybridization; ROC, receiver operating characteristic; AUC, area under curve; TCIA, The Cancer Imaging Archive; GTV, gross tumor volume; IBSI, International Biomarker Standardization Initiative; RQS, radiomics quality score; TRIPOD, Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis; CUI, clinical utility index; GLCM, gray level co-occurrence; GLRLM, gray level run-length; GLSZM, gray level size-zone texture matrices.

* Corresponding author.
E-mail address: roberto.casale25@gmail.com (R. Casale).

[1] Both authors contributed equally.

https://doi.org/10.1016/j.ejrad.2021.109678
Received 12 December 2020; Received in revised form 4 March 2021; Accepted 21 March 2021
Available online 5 April 2021
0720-048X/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

years [3]. LGG finally advances to higher grades, with a significantly lower survival rate [3].

Treatment choices for LGG are based on WHO grades, molecular profiles, and patient characteristics (e.g. age and Karnofsky performance status) [4]. The co-deletion of chromosome arms 1p and 19q has an important role in choosing the right treatment, indeed co-deletion is a useful prognostic molecular marker as it can be used for the prediction of response to radiotherapy and chemotherapy, and it is associated with longer survival [5–8]. Thus, efficient treatment planning necessitates proper classification of WHO grade and 1p/19q co-deletion status.

The 1p/19q status can be determined by different techniques: fluorescence in situ hybridization (FISH), polymerase chain reaction, array comparative genomic hybridization, or multiplex ligation-dependent probe amplification [9]. This molecular classification is achieved through histopathologic examination; albeit being the reference standard for this task, it has some limits, such as limited surgical accessibility and heterogeneity of the sampled tissue. Furthermore, biopsy samples are not representative of the whole neoplasm [10].

The unmet clinical need is to find a non-invasive and robust classification method of 1p/19q status of the entire tumor volume in order to effectively direct treatment planning of LGG [5–8] for cases when complete resection cannot be performed and/ or where the biopsy cannot be obtained from the tumor. Most notably, in childhood tumors around 30%–50% of LGGs are inoperable as a result of their position in highly eloquent areas of the brain [11]. Currently, MRI is a useful technique in order to obtain helpful data for therapy decisions, and for pre-therapeutic noninvasive diagnosis.

Radiomics is a research field whose scope is to extract imaging features from radiographic images (including MRI) that can potentially capture phenotypic, genomic, proteomics patterns having prognostic value and clinical significance. The underlying hypothesis of radiomics is that medical imaging may express additional data correlating with genomic and proteomics patterns and can be manifested in macroscopic image-based features, not visible by the unaided eye and thus not used [12–14].

In the last few years different studies have demonstrated that 1p19q status can be predicted using MRI [15–22]. Furthermore, Branzoli et al. [23] recently identified elevated levels of cystathionine in 1p/19q codeleted gliomas compared to non-codeleted gliomas, using in vivo magnetic resonance spectroscopy. In our analysis, routine MRI sequences were used, without additional experimental or expensive MRI sequences.

The main purpose of this study was to develop and to validate a non-invasive method to predict the 1p/19q status of LGG from T2-weighted and T1-weighted post-contrast MRI images using texture analysis as an alternative to surgical biopsy. The secondary aim was to compare two voxel resampling methods: radiomics features calculated from images resampled using cubic and linear interpolation methods. Cubic spline and convolution interpolation are third-order methods that typically interpolate smoother surfaces than linear methods, while they are known to be slower in implementation [24]. Linear interpolation is a commonly used algorithm since it is computationally cheap and leads neither to rough blocking artifacts images that are generated by nearest neighbor techniques, nor will it cause out-of-range gray levels that might be produced by higher order interpolation [25].

2. Materials and methods

2.1. Data

The training dataset consisted of 159 LGG patients with pre-operative MRI images and 1p/19q status proven by biopsy. They were identified within the LGG-1p19q Deletion dataset [15,26,27] on The Cancer Imaging Archive (TCIA). The validation dataset consisted of similar patient data of 50 randomly selected patients, also from TCIA, albeit in the TCGA-LGG dataset [27,28]. For TCGA, the 1p/19q status for validation dataset was derived from a previous study based on this dataset [16]. Patients were selected according to the following inclusion criteria: exams with a slice thickness $\leq$ 7.5 mm, artifacts in less than 50 % of the slices containing the gross tumor volume (GTV) visually assessed by a radiologist with 3 years' experience (R.C.), and the presence of T2-weighted and contrast enhanced T1-weighted images and 1p/19q status. The GTV was delineated using MIM software version 6.9.0 (MIM, Cleveland, United States) by one radiologist (R.C.).

2.2. Image pre-processing and radiomics feature extraction

In order to somewhat account for inter-scanner variability, Z-score normalization was applied to the GTVs in each image series (per patient). The formula for Z-score normalization for GTV intensities is:

$$\frac{original\ intensity\ value\ -\ \mu}{\sigma}$$

where μ is the mean intensity inside each GTV and σ is the intensity standard deviation in each GTV.

Voxel size resampling was performed before feature extraction using cubic and linear interpolation separately. Images were resampled to a voxel size of $3 \times 3 \times 3\ mm^3$; more information about the choice of voxel size can be found in the Supplementary Materials (Voxel size section).

To reduce noise and computational burden, grayscale values were aggregated into the same number of bins (50 bins) for all MRI exams. The fixed bin number method was used to achieve a better normalizing effect as intensity units are not absolute in MRI [25]. Radiomics features compliant with the International Biomarker Standardization Initiative (IBSI), as well as non-IBSI features were extracted with the RadiomiX research software (supported by Oncoradiomics, Liège, Belgium).

Radiomics features were extracted consisting of five main groups: 1) fractal features, 2) first order statistics, 3) shape and size, 4) texture descriptors including gray level co-occurrence (GLCM), gray level run-length (GLRLM) and gray level size-zone texture matrices (GLSZM), 5) features from groups 1, 3 and 4 after wavelet decomposition. GLCM distance was 1. Definitions and detailed feature list are described in Supplementary Materials (Tables 3 and 4).

2.3. Feature selection and statistical analysis

Fig. 1 illustrates the 3 steps that were performed only on the training dataset for feature selection; the second step was repeated 300 times, with different sample groupings. All this procedure was performed twice, for cubic and linear interpolation respectively. The first step used correlation-based feature subset selection (CfsSubsetEval function, Weka software version 3.8.3) [29–31] to eliminate irrelevant and redundant features. In the second step, a table was created that ordered and ranked features according to their importance using a 10-fold cross validation treebag recursive feature elimination algorithm (RFE) (Python 3.7.6 version, scikit-learn 0.21.2 package). Finally, in the third step, a learning curve was computed (AUC vs. the incremental number of features obtained from the ranked feature table). More information about the 3 steps are explained in Supplementary Materials (Feature selection section).

Inter-correlation among selected features and with volume were calculated with the Spearman correlation coefficient. Moreover, the Mann-Whitney test was applied in order to check statistically significant differences in GTV values in codeleted/non co-deleted groups in the training dataset.

Statistical analysis was performed with Python 3.7.6 version (scipy 1.4.1 package, pandas 1.0.0 package).

2.4. Classification

A random forest (RF) classification model was trained on the training

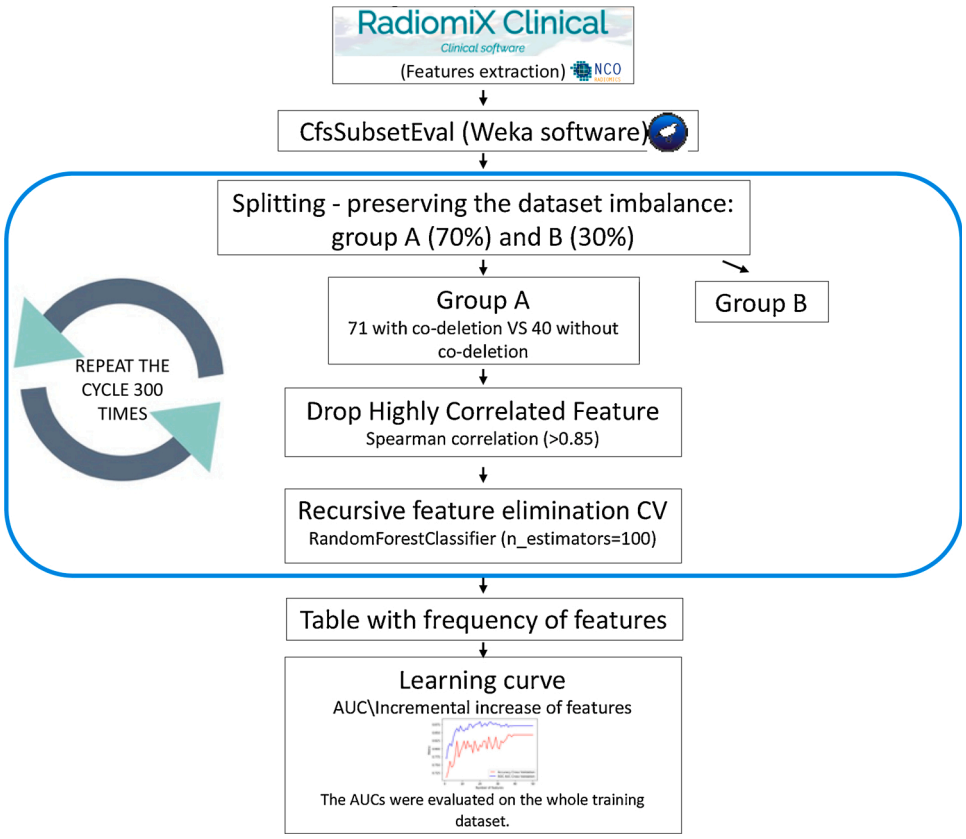

Fig. 1. Feature selection (only training dataset).

dataset with the selected features, and performance metrics calculated when applied to both datasets without further adjustments. To mitigate the effect of the unbalanced outcomes, the training dataset was balanced using an adaptive synthetic (ADASYN) resampling approach, which creates artificial patients for the minority class, before the RF model was trained [32]. On the training dataset, internal 10-fold cross-validation was performed, followed by a bootstrap method (n = 10000) to have an evaluation of the error of the performance metrics (median, 2.5 and 97.5 percentiles). On the validation dataset, a bootstrap method (n = 10000) was implemented, and the median values and 2.5 and 97.5 percentiles calculated. During the cross-validation procedure, each set preserved roughly the same ratio of samples for each class (co-deleted/non co-deleted) as the complete training dataset and ADASYN applied to the training fold.

Accuracy, sensitivity, specificity, receiver operating characteristic curve (ROC) and AUC were computed. All these steps of the workflow were repeated twice (for cubic interpolation and linear interpolation). Classification performance was compared for cubic and linear interpolation-based data for both training dataset cross-validation results and validation results; the De-Long test was used to compare AUCs obtained from each model.

This segment and statistical analysis were performed with Python 3.7.6 version (scikit-learn 0.21.2 package, scipy 1.4.1 package), and R 3.6.1 version (pROC 1.14.0 package).

2.5. TRIPOD and Radiomics quality score

This study followed the instruction of Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD), and the Radiomics quality score (RQS) was used to evaluate the radiomics workflow [13,33,34]. The RQS score for this specific study was 44 %. The RQS maximum score is 100 % and it is based on a 36 points system; a high value reveals a higher methodological quality research and reporting [33].

2.6. Clinical utility index (CUI)

Clinical utility indexes were computed for the RF model tested on external validation dataset. CUI was developed in 2007 and aimed to take into account both occurrence and discrimination [35–40]. The value for CUI ranges from 0 to 1: excellent utility (CUI ≥ 0.81), good utility (CUI ≥ 0.64), satisfactory/fair utility (CUI ≥ 0.49), poor utility (CUI ≤ 0.49) and very poor utility (CUI ≤ 0.36) [36]. More information and relative formulas about CUI are reported in Supplementary Materials (Clinical utility index section).

2.7. Data sharing

The dataset and GTV used in this article can be provided upon contact with the corresponding author.

The python code used for the feature selection, classification model and evaluation of the algorithm is available on GitHub https://github.com/roberto-casale/LGG-1p-19q-deletion.

3. Results

3.1. Data

3.1.1. Training dataset

One hundred and fifty-nine consecutive LGG patients with preoperative MRI images collected between 01-10-2002 and 01-09-2011 and biopsy proven 1p/19q status were identified within the LGG-1p19q Deletion archive (Supplementary Materials Table 1). The data included 102 patients with co-deleted 1p/19q arms and 57 with non-co-deleted arms. The grades of the LGG lesions were II (n = 104) and III (n = 55). The types of LGG were oligoastrocytoma (n = 97),

oligodendrogliomas (n = 45), and astrocytomas (n = 17). Median age was 42 (range 13–84) and this dataset included 76 women and 83 men. Post-contrast T1- and T2-weighted images were available for all selected patients. All images were acquired with 1.5 or 3 T scanners, slice thicknesses ranged from 1 to 7.5 mm and isotropic pixel size in the axial plane ranged from 0.43 to 1.09 mm. More details about exams are shown in Supplementary Materials and Supplementary Tables.

3.1.2. External validation dataset

Fifty (n = 50) patients were randomly chosen from the TCGA-LGG dataset (Supplementary Materials Table 2) while maintaining outcome balance. The 1p/19q status was identified thanks to Supplementary Tables from the research of Chia Feng Lu al. [15] that used the same dataset. This validation dataset included 25 non-deleted and 25 co-deleted LGG. The grades of LGG were II (n = 29) and III (n = 21). The types of LGG were oligoastrocytoma (n = 14), oligodendrogliomas (n = 28), and astrocytomas (n = 8). Median age was 46 (range 20–74) and it included 22 women and 28 men. Post-contrast T1- and T2-weighted images were available for all selected patients. All images were acquired with 1.5 T or 3 T scanners (it was not reported the magnetic field for five patients), and the slice thickness ranged from 0.9 mm to 7.5 mm and isotropic pixel size in the axial plane ranged from 0.39 to 1.02 mm. More details about exams are shown in Supplementary Materials and Supplementary Tables.

No significant differences in gender (M/F = 1.1 in training set vs M/F = 1.3 in validation set) and WHO grade ratios (II/III = 1.9 in training set vs II/III = 1.4) were observed between the training and validation sets. There were significant differences in histology and age (mean age 46.5 in training set vs 41.6 in validation set). Level of significance was $\alpha = 0.05$ for Chi-square tests and Mann-Whitney test (age comparison). Demographic and clinical data description are presented in Table 1.

3.2. Radiomics features extraction, selection, and statistical analysis

In total, 5352 radiomics features per patient were extracted from both T1- and T2- weighted images; 2676 features extracted with each of cubic and linear interpolation voxel resampling methods.

After correlation-based feature subset selection a total of 48 features remained for cubic interpolation and 51 features for linear interpolation.

These remaining features were fed into 300 loops of 10-fold cross validation RFE. Supplementary Materials Table 5 (for cubic interpolation) and Supplementary Materials Table 6 (for linear interpolation) show how many times each feature was selected during the 300 loops. GTV volume was not chosen among the selected features.

With these ranked features, two learning curves were computed (AUC vs. incremental increase of features) respectively for cubic interpolation (Supplementary Materials Fig. 1) and linear interpolation (Supplementary Materials Fig. 2) using only the training dataset. The classifier used to generate the curve was RF, with co-deleted/ non co-deleted outcome and 10- fold cross validation.

The number of features for the final model was chosen near the first salient point of the learning curve for AUC score. All features that went into the model satisfy the condition that they were selected more than 68 % (greater than or equal to 205 times) in the RFE loops to ensure a certain level of robustness.

Finally, the selected features were 7 for cubic interpolation (Table 2) and 5 for linear interpolation (Table 3).

The inter-correlation among the selected features in the training dataset is shown in Supplementary Materials Fig. 3 (cubic interpolation) and Supplementary Materials Figure 4 (linear interpolation).

The inter-correlation among the selected features in external validation dataset is shown in Supplementary Materials Figure 5 (cubic interpolation) and Supplementary Materials Figure 6 (linear interpolation). These results were computed with Spearman rank correlation.

Value distributions for selected features for codeletion and non-codeletion classes in cases of cubic and linear interpolation are presented in Supplementary Materials Figure 7 and Figure 8, respectively.

3.3. Classification

3.3.1. Results on training dataset

All results are reported as the median [2.5 percentile – 97.5 percentile]. For cubic interpolation, the RF model achieved an AUC of 0.86 [0.81−0.91] and for linear interpolation an AUC of 0.82 [0.75−0.87] (Table 4).

The De-Long test was used to compare model performances obtained from models trained on data that underwent cubic and linear interpolation. According to the results of this test, there was no statistically significant difference between the two AUCs (p-value = 0.073).

The Mann-Whitney, applied to GTV values in codeleted/non co-deleted groups, shows no statistical difference between the two groups (p = 0.149; alpha = 0.05).

3.3.2. Results on validation dataset

The AUC for features extracted for cubic interpolation was 0.87 [0.76−0.95] and for linear interpolation was 0.77 [0.61−0.89] (Table 5). According to the DeLong test there was no statistically significant difference between the two models (p-value = 0.178).

The confusion matrix for the two models are shown in Supplementary Materials Table 7 (for cubic interpolation) and Table 8 (for linear interpolation).

The ROC curves are shown in Fig. 2 (for cubic interpolation) and Fig. 3 (for linear interpolation).

3.4. Clinical utility index (CUI)

The positive CUI, calculated for the RF model with cubic interpolation features and tested on validation dataset, was 0.451 (CI: 0.203−0.698); the negative CUI was 0.605 (CI: 0.483−0.762). The positive and negative CUI values obtained with cubic interpolation features had a poor and a satisfactory/fair utility value respectively.

The positive CUI for the RF model, obtained with linear interpolation

Table 1
Data description.

	Training dataset	Validation dataset	p-value*
Number of patients	159	60	–
Age, y, mean (SD)	41.6 (13.8)	46.5 (13.0)	0.026
Gender ratio (M/F)	83/76	28/22	0.759
Grade ratio (II/III)	104/55	29/21	0.435
Histology ratio (astrocytoma/ oligoastrocytoma/ oligodendroglioma)	17/97/45	8/14/28	0.000
Outcome ratio (codeletion/non-codeletion)	102/57	25/25	–

* p-value for statistically significant differences of value distribution in training and validation datasets: age – Mann-Whitney, gender ratio – chi-square, grade ratio-chi-square, histology ratio – chi square.

Table 2
Selected features for cubic interpolation with frequency.

Features	Frequency*
GLCM_average (T2)	293
Wavelet_LHL_Stats_median (T1)	289
Wavelet_LLH_Stats_median (T1)	276
GLCM_clusShade (T1)	258
Wavelet_LHH_Fractal_lacunarity (T2)	223
Wavelet_HLL_GLCM_correl1 (T2)	212
Wavelet_LLL_Stats_p10 (T2)	206

* Frequency is number of times the feature was selected during the 300 loops.

Table 3
Selected features for linear interpolation with frequency.

Features	Frequency*
Wavelet_LLH_Stats_median (T1)	275
Wavelet_LHL_Stats_median (T1)	271
Wavelet_LLL_IH_p10 (T1)	257
GLCM_clusShade (T1)	213
Wavelet_LHH_Fractal_lacunarity (T2)	205

* Frequency is number of times the feature was selected during the 300 loops.

Table 4
Classification performance on training dataset (10- fold cross validation).

	Cubic interpolation median [2.5–97.5 percentile]	Linear interpolation median [2.5–97.5 percentile]
Accuracy	0.81[0.75−0.86]	0.76 [0.71−0.82]
Sensitivity	0.77 [0.69−0.85]	0.72 [0.63−0.8]
Specificity	0.85 [0.78−0.92]	0.81 [0.74−0.88]
AUC	0.86 [0.81−0.91]	0.82 [0.75−0.87]

Table 5
Classification performance on external validation dataset.

	Cubic interpolation [2.5-percentile - 97.5 percentile]	Linear interpolation [2.5-percentile - 97.5 percentile]
Accuracy	0.72 [0.6−0.82]	0.72 [0.6−0.84]
Sensitivity	0.52 [0.32−0.72]	0.6 [0.4−0.8]
Specificity	0.92 [0.8–1.0]	0.84 [0.68−0.96]
AUC	0.87 [0.76−0.95]	0.77 [0.61−0.89]

features and tested on validation dataset, was 0.474 (CI: 0.238−0.709), so with a poor utility value; the negative CUI for the model obtained with linear interpolation features was 0.569 (CI: 0.435−0.703), so with a satisfactory/fair utility value.

These results showed that the RF model, trained both with cubic and linear interpolation features, achieved a satisfactory negative CUI, meaning that this algorithm can be reasonably useful for screening patients with 1p-19q non-co-deletion status. On the other side, the RF model, trained both with cubic and linear interpolation features, achieved a poor positive CUI, meaning that this method has low utility to confirm patients with non-co-deleted status; in practical terms, if a patient obtains a result that suggest having non-co-deleted status, this patient should be further studied to confirm the non-co-deleted status.

4. Discussion

In this study we explored the ability of radiomics features extracted from the GTV on preoperative MRI (acquired with T1-weighted contrast enhanced and T2-weighted sequences) to predict molecular status of chromosome 1p/19q co-deletion in LGG patients. To investigate the influence of the resampling method on the classification models' performance, we used both cubic and linear interpolation kernels for further comparison. After feature selection, the feature vectors contained 5 and 7 features for cubic and linear interpolation-based data, respectively. These feature vectors only had 4 common features (Wavelet_LLH_Stats_median (T1), Wavelet_LHL_Stats_median (T1), GLCM_clusShade (T1) and Wavelet_LHH_Fractal_lacunarity (T2)). Therefore, we conclude that the method presented is not completely robust to the resampling method and additional studies on features reproducibility are needed.

According to Spearman correlation coefficient, within the training dataset, the feature vectors consisted of statistically independent

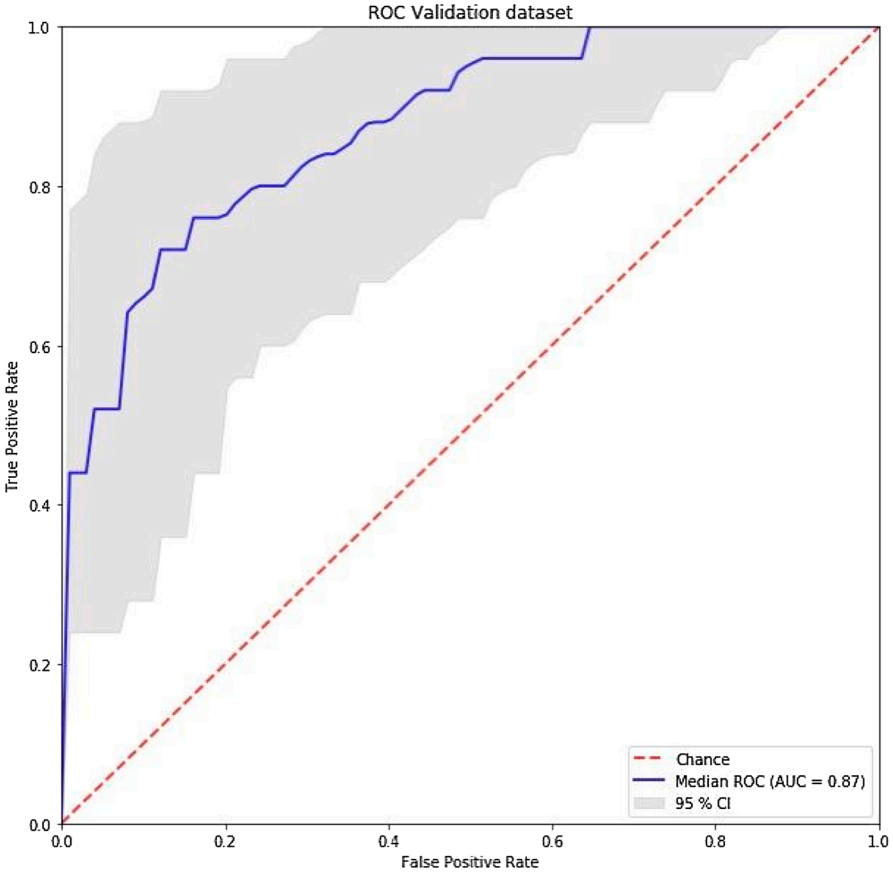

Fig. 2. ROC and AUC for features extracted with cubic interpolation - Results on validation dataset.

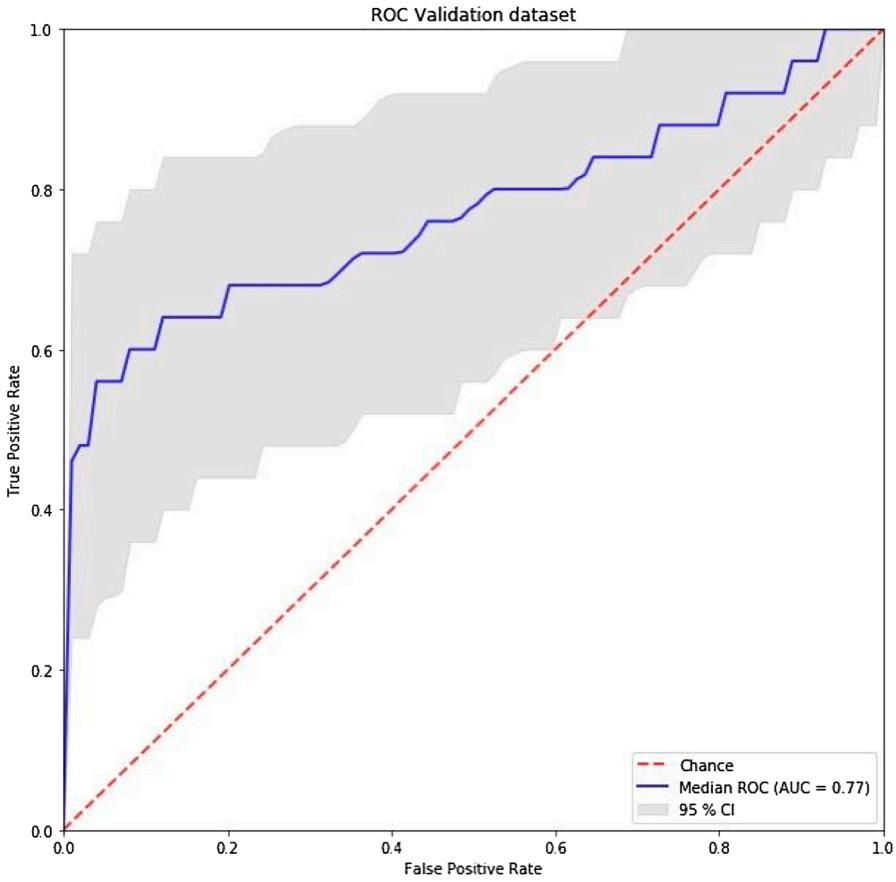

Fig. 3. ROC and AUC for features extracted with linear interpolation - Results on validation dataset.

features. In validation dataset, some of these features are correlated to each other (Spearman correlation coefficient 0.76 and 0.77 for T1-weighted Wavelet_LHL_Stats_median (T1) and Wavelet_LLH_Stats_median (T1) for cubic and linear interpolation-based features, respectively).

According to De-Long test, there were no statistically significant differences between AUCs obtained from the cubic interpolation model and linear interpolation model both on training dataset (p-value = 0.178) and validation dataset (p-value = 0.073).

The advantages of the present study are its non-invasiveness, the analysis of the entire volume of the lesion, and the ubiquitous availability, as it is based on simple conventional MRI sequences.

Other studies try to predict 1p/19q status, some of which aim to solve the same problem using MRI [15–22]. They all are using multimodal conventional MRI data, most often combining T2-weighted and contrast enhanced T1-weighted data together. Classification performance of the present study did not exceed the results, obtained in [15–17,19,20]. Nevertheless, the present study has some benefits over previously mentioned studies [15–17,20]: (1) the potential reproducibility, achieved with open source data usage and utilization of an automated pipeline, (2) the potential interpretability of results, as input features are known and understood, (3) the presence of clinical utility evaluation, (4) the evaluation of two different resampling methods.

The present study has some limitations. The main limitation is the relatively small sample size, which decreases statistical power of the classification results. For this reason, to test the model, we performed cross-validation on the training dataset and then we trained it on the whole training dataset to perform validation on external dataset. Also, for this reason, to estimate model performance and its error on external validation dataset, we performed a bootstrapping approach, which produces multiple instances of the same observations and omits other original observations. The second limitation was related to data balance within and between training and validation datasets. Outcomes in the training dataset were significantly unbalanced (102 cases of codeletion vs 57 cases of non-codeletion); to partially overcome this limitation, the ADASYN method was used, which is not without uncertainties. The third limitation was related to significant differences in histology and age distribution in training and validation datasets. Histology effect and age have not been investigated and included into models and they could be explored in further studies. The fourth limitation was related to different MRI field strengths, values of slice thickness (0.9–7.5 mm) and isotropic pixel spacing (0.39–1.09 mm); these differences could be a source of batch effects, modifying radiomics features significantly, but also could be an opportunity to test the stability of methods across different image acquisition parameters. The fifth limitation arises from possible bias stemming from the random selection for 50 patients inside the validation dataset.

In summary, the proposed non-invasive method is able to predict molecular status of chromosome 1p/19q co-deletion in LGG patients, based on multi-scanner multi-field MRI data. Although there is still room for improvement in accuracy metrics, its usefulness was indicated for the estimation of prognostic molecular markers. Results of its validation on external data demonstrated its generalizability. According to the results of statistical tests, there were no statistically significant differences between the AUCs obtained with different spatial resampling interpolation methods (cubic and linear).

Regarding the diagnostic utility of this method, the CUI demonstrated that the RF model (trained both with cubic and linear interpolation features) achieved a satisfactory negative CUI, while the RF model (trained both with cubic and linear interpolation features) achieved a poor positive CUI. Therefore, linear and cubic models can be reasonably helpful for ruling out non-co-deleted status, but they can be poorly

useful for confirming non-co-deleted status. This difference can be explained by the different accuracy metrics: indeed, both algorithms had specificity and positive predictive values higher than sensitivity and negative predictive values; moreover, the unbalanced class in the training dataset could affect the performance. These results should be considered in future studies and should be taken into account in a future clinical scenario.

This approach may be an opportunity to help medical decision. Despite the dataset was limited, ADASYN increased the number of cases in the training phase. However, further studies based on more heterogeneous and larger patient population are mandatory to confirm and validate our current results.

5. Conclusions

MRI radiomics analysis, based on T2-weighted and T1-weighted post-contrast images, could supply a reliable noninvasive technique for the prediction of 1p/19q status in LGGs, giving useful information for personalized therapy assessment and pretreatment prediction. Regarding the two different voxel resampling methods, no statistically significant differences were found.

CRediT authorship contribution statement

Roberto Casale: Conceptualization, Methodology, Software, Formal analysis, Resources, Writing - original draft. **Elizaveta Lavrova:** Methodology, Software, Formal analysis, Investigation. **Sebastian Sanduleanu:** Methodology, Visualization, Formal analysis, Writing - review & editing. **Henry C. Woodruff:** Supervision, Methodology. **Philippe Lambin:** Supervision, Funding acquisition.

Declaration of Competing Interest

P.L. reports—within and outside the submitted work—grants or sponsored research agreements from Varian Medical, Oncoradiomics, ptTheragnostic/DNAmito, and Health Innovation Ventures. He received an advisor/presenter fee and/or reimbursements of travel costs/external grant writing fee and/or in-kind manpower contribution from Oncoradiomics, BHV, Merck, Varian, Elekta, ptTheragnostic and Convert Pharmaceuticals. P.L. has minority shares in the company Oncoradiomics, Convert Pharmaceuticals, The Medical Cloud Company and LivingMed Biotech, and is co-inventor of two issues patents with royalties on radiomics (PCT/NL2014/050248, PCT/NL2014/050728) licensed to Oncoradiomics, one issue patent on mtDNA (PCT/EP2014/059089) licensed to ptTheragnostic/DNAmito, three non-patented inventions (software) licensed to ptTheragnostic/DNAmito and Oncoradiomics and Health Innovation Ventures, and three non-issues, non-licensed patents on Deep Learning-Radiomics and LSRT (N2024482, N2024889, N2024889). H.W. reports minority shares in OncoRadiomics.

Acknowledgements

Authors acknowledge financial support from ERC advanced grant (ERC-ADG-2015 n° 694812 - Hypoximmuno), ERC-2018-PoC: 813200-CL-IO, ERC-2020-PoC: 957565-AUTO.DISTINCT, Authors also acknowledge financial support from EUROSTARS (DART, DECIDE), the European Union's Horizon 2020 research and innovation programme under grant agreement: ImmunoSABR n° 733008, MSCA-ITN-PREDICT n° 766276, FETOPEN- SCANnTREAT n° 899549, CHAIMELEON n° 952172, EuCanImage n° 952103, TRANSCAN Joint Transnational Call 2016 (JTC2016 CLEARLY n° UM 2017-8295) and Interreg V-A Euregio Meuse-Rhine (EURADIOMICS n° EMR4).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ejrad.2021.109678.

References

[1] S. Cha, Update on brain tumor imaging: from anatomy to physiology, AJNR Am. J. Neuroradiol. (2006) 475–487.
[2] A. Lanese, E. Franceschi, A.A. Brandes, The risk assessment in low-grade gliomas: an analysis of the european organization for research and treatment of Cancer (EORTC) and the radiation therapy oncology group (RTOG) criteria, Oncol. Ther. 6 (2) (2018) 105–108.
[3] E.B. Claus, K.M. Walsh, J.K. Wiencke, A.M. Molinaro, J.L. Wiemels, J. M. Schildkraut, M.L. Bondy, M. Berger, R. Jenkins, M. Wrensch, Survival and low-grade glioma: the emergence of genetic information. Neurosurgical Focus, American Association of Neurological Surgeons, 2015.
[4] A. Picca, G. Berzero, M. Sanson, Current Therapeutic Approaches to Diffuse Grade II and III Gliomas, Therapeutic Advances in Neurological Disorders, SAGE PublicationsSage UK, London, England, 2018.
[5] S. Fellah, D. Caudal, A.M. De Paula, P. Dory-Lautrec, D. Figarella-Branger, O. Chinot, P. Metellus, P.J. Cozzone, S. Confort-Gouny, B. Ghattas, V. Callot, N. Girard, Multimodal MR imaging (Diffusion, perfusion, and spectroscopy): is it possible to distinguish oligodendroglial tumor grade and 1p/19q codeletion in the pretherapeutic diagnosis? Am. J. Neuroradiol. (2013) 1326–1333.
[6] N.L. Jansen, C. Schwartz, V. Graute, S. Eigenbrod, J. Lutz, R. Egensperger, G. Pöpperl, H.A. Kretzschmar, P. Cumming, P. Bartenstein, J.-C. Tonn, F.-W. Kreth, C. la Fougère, N. Thon, Prediction of oligodendroglial histology and LOH 1p/19q using dynamic [(18)F]FET-PET imaging in intracranial WHO grade II and III gliomas, Neurooncology (2012) 1473–1480. Oxford University Press.
[7] Y. Iwadate, N. Shinozaki, T. Matsutani, Y. Uchino, N. Saeki, Molecular imaging of 1p/19q deletion in oligodendroglial tumours with 11C-methionine positron emission tomography, J. Neurol. Neurosurg. Psychiatr. (2016) 1016–1021. BMJ Publishing Group.
[8] P. Bourdillon, C. Hlaihel, J. Guyotat, L. Guillotton, J. Honnorat, F. Ducray, F. Cotton, Prediction of anaplastic transformation in low-grade oligodendrogliomas based on magnetic resonance spectroscopy and 1p/19q codeletion status, J. Neurooncol. (2015) 529–537. Springer US.
[9] A. Woehrer, P. Sander, C. Haberler, S. Kern, H. Maier, M. Preusser, C. Hartmann, J. M. Kros, J.A. Hainfellner, FISH-based detection of 1p 19q codeletion in oligodendroglial tumors: procedures and protocols for neuropathological practice - A publication under the auspices of the Research Committee of the European Confederation of Neuropathological Societies (Euro-CNS), Clin. Neuropathol. (2011) 47–55.
[10] H. Sasaki, M.C. Zlatescu, R.A. Betensky, L.B. Johnk, A.N. Cutone, J.G. Cairncross, D.N. Louis, Histopathological-molecular genetic correlations in referral pathologist-diagnosed low-grade "oligodendroglioma", J. Neuropathol. Exp. Neurol. (2002) 58–63. Narnia.
[11] M.I. Ruge, T. Simon, B. Suchorska, R. Lehrke, C. Hamisch, F. Koerber, M. Maarouf, H. Treuer, F. Berthold, V. Sturm, J. Voges, Stereotactic brachytherapy with iodine-125 seeds for the treatment of inoperable low-grade gliomas in children: long-term outcome, J. Clin. Oncol. 29 (31) (2011) 4151–4159.
[12] H.J. Aerts, E.R. Velazquez, R.T. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, F. Hoebers, M. M. Rietbergen, C.R. Leemans, A. Dekker, J. Quackenbush, R.J. Gillies, P. Lambin, Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach, Nat. Commun. 5 (2014) 4006.
[13] P. Lambin, R.T.H. Leijenaar, T.M. Deist, J. Peerlings, E.E.C. de Jong, J. van Timmeren, S. Sanduleanu, R. Larue, A.J.G. Even, A. Jochems, Y. van Wijk, H. Woodruff, J. van Soest, T. Lustberg, E. Roelofs, W. van Elmpt, A. Dekker, F. M. Mottaghy, J.E. Wildberger, S. Walsh, Radiomics: the bridge between medical imaging and personalized medicine, Nat. Rev. Clin. Oncol. 14 (12) (2017) 749–762.
[14] P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R.G. van Stiphout, P. Granton, C.M. Zegers, R. Gillies, R. Boellard, A. Dekker, H.J. Aerts, Radiomics: extracting more information from medical images using advanced feature analysis, Eur. J. Cancer 48 (4) (2012) 441–446.
[15] Z. Akkus, I. Ali, J. Sedlář, J.P. Agrawal, I.F. Parney, C. Giannini, B.J. Erickson, Predicting deletion of chromosomal arms 1p/19q in low-grade gliomas from MR images using machine intelligence, J. Digit. Imaging (2017) 469–476.
[16] C.-F. Lu, F.-T. Hsu, K.L.-C. Hsieh, Y.-C.J. Kao, S.-J. Cheng, J.B.-K. Hsu, P.-H. Tsai, R.-J. Chen, C.-C. Huang, Y. Yen, C.-Y. Chen, Machine learning–Based radiomics for molecular subtyping of gliomas, Clin. Cancer Res. (2018) 4429–4436.
[17] S. Zhang, G.C.-Y. Chiang, R.S. Magge, H.A. Fine, R. Ramakrishna, E.W. Chang, T. Pulisetty, Y. Wang, W. Zhu, I. Kovanlikaya, MRI based texture analysis to classify low grade gliomas into astrocytoma and 1p/19q codeleted oligodendroglioma, Magn. Reson. Imaging (2019) 254–258.
[18] S.R. van der Voort, F. Incekara, M.M.J. Wijnenga, G. Kapas, M. Gardeniers, J. W. Schouten, M.P.A. Starmans, R. Nandoe Tewarie, G.J. Lycklama, P.J. French, H. J. Dubbink, M.J. van den Bent, A. Vincent, W.J. Niessen, S. Klein, M. Smits, Predicting the 1p/19q codeletion status of presumed low-grade glioma with an externally validated machine learning algorithm, Clin. Cancer Res. 25 (24) (2019) 7455–7462.

[19] B. Kocak, E.S. Durmaz, E. Ates, I. Sel, S. Turgut Gunes, O.K. Kaya, A. Zeynalova, O. Kilickesmez, Radiogenomics of lower-grade gliomas: machine learning-based MRI texture analysis for predicting 1p/19q codeletion status, Eur. Radiol. 30 (2) (2020) 877–886.
[20] Z. Kong, C. Jiang, Y. Zhang, S. Liu, D. Liu, Z. Liu, W. Chen, P. Liu, T. Yang, Y. Lyu, D. Zhao, H. You, Y. Wang, W. Ma, F. Feng, Thin-slice magnetic resonance imaging-based radiomics signature predicts chromosomal 1p/19q Co-deletion status in grade II and III gliomas, Front. Neurol. 11 (2020) 551771.
[21] Z.A. Shboul, J. Chen, M.I. K, Prediction of molecular mutations in diffuse low-grade gliomas using MR imaging features, Sci. Rep. 10 (1) (2020) 3711.
[22] A.P. Bhandari, R. Liong, J. Koppen, S.V. Murthy, A. Lasocki, Noninvasive determination of IDH and 1p19q status of lower-grade gliomas using MRI radiomics: a systematic review, AJNR Am. J. Neuroradiol. 42 (1) (2021) 94–101.
[23] F. Branzoli, C. Pontoizeau, L. Tchara, A.L. Di Stefano, A. Kamoun, D.K. Deelchand, R. Valabrègue, S. Lehéricy, M. Sanson, C. Ottolenghi, M. Marjańska, Cystathionine as a marker for 1p/19q codeleted gliomas by in vivo magnetic resonance spectroscopy, NeuroOncology (2019) 765–774.
[24] L. Ruijiang, X. Lei, N. Sandy, L.R. Daniel, Radiomics and Radiogenomics: Technical Basis and Clinical Applications, Chapman and Hall/CRC, 2019.
[25] A. Zwanenburg, S. Leger, M. Vallières, S. Löck, Image biomarker standardisation initiative, arXiv 1612 (2019), 07003v11.
[26] B. Erickson, Z. Akkus, J. Sedlar, P. Korfiatis, Data from LGG-1p19qDeletion. The Cancer Imaging Archive, 2017.
[27] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, F. Prior, The cancer imaging archive (TCIA): maintaining and operating a public information repository. J. Digit. Imaging, 2013, pp. 1045–1057.
[28] N. Pedano, A.E. Flanders, L. Scarpace, T. Mikkelsen, J.M. Eschbacher, B. Hermes, V. Sisneros, J. Barnholtz-Sloan, Q. Ostrom, Radiology data from the Cancer genome atlas low grade glioma [TCGA-LGG] collection. The Cancer Imaging Archive, 2016.
[29] E. Frank, M.A. Hall, I.H. Witten, The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", fourth edition, Morgan Kaufmann, 2016.
[30] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The WEKA Data Mining Software, ACM SIGKDD Explorations Newsletter, 2009, p. 10.
[31] M.A. Hall, Correlation-based Feature Subset Selection for Machine Learning, Hamilton, New Zealand, 1999.
[32] H. Haibo, B. Yang, A.G. Edwardo, L. Shutao, ADASYN: adaptive synthetic sampling approach for imbalanced learning, in: IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, 2008, pp. 1322–1328.
[33] S. Sanduleanu, H.C. Woodruff, E.E.C. de Jong, J.E. van Timmeren, A. Jochems, L. Dubois, P. Lambin, Tracking tumor biology with radiomics: a systematic review utilizing a radiomics quality score, Radiother. Oncol. 127 (3) (2018) 349–360.
[34] K.G. Moons, D.G. Altman, J.B. Reitsma, J.P. Ioannidis, P. Macaskill, E. W. Steyerberg, A.J. Vickers, D.F. Ransohoff, G.S. Collins, Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration, Ann. Intern. Med. 162 (1) (2015) W1–73.
[35] A.J. Mitchell, The clinical significance of subjective memory complaints in the diagnosis of mild cognitive impairment and dementia: a meta-analysis, Int. J. Geriatr. Psychiatry 23 (11) (2008) 1191–1202.
[36] M. Pentzek, A. Wollny, B. Wiese, F. Jessen, F. Haller, W. Maier, S.G. Riedel-Heller, M.C. Angermeyer, H. Bickel, E. Mosch, S. Weyerer, J. Werle, C. Bachmann, T. Zimmermann, H. van den Bussche, H.H. Abholz, A. Fuchs, G. AgeCoDe Study, Apart from nihilism and stigma: what influences general practitioners' accuracy in identifying incident dementia? Am. J. Geriatr. Psychiatry 17 (11) (2009) 965–975.
[37] D.C. Goncalves, E. Arnold, K. Appadurai, G.J. Byrne, Case finding in dementia: comparative utility of three brief instruments in the memory clinic setting, Int. Psychogeriatr. 23 (5) (2011) 788–796.
[38] A.J. Mitchell, How do we know when a screening test is clinically useful? in: A. J. Mitchell, C. James (Eds.), Screening for Depression in Clinical Practice: an Evidence-Based Guide Coyne, 2009. ISBN10: 0195380193 OUP.
[39] R.R. Davies, A.J. Larner, Addenbrooke's cognitive examination (ACE) and its revision (ACE-R), in: A.J. Larner (Ed.), Cognitive Screening Instruments: A Practical Approach, Springer, London, London, 2013, pp. 61–77.
[40] A.J. Mitchell, Sensitivity x PPV is a recognized test called the clinical utility index (CUI+), Eur. J. Epidemiol. 26 (3) (2011) 251–252, author reply 252.