Clinical and Translational Radiation Oncology 7 (2017) 49–54

Contents lists available at ScienceDirect

Clinical and Translational Radiation Oncology

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

Original Research Article

Radiomic analysis in prediction of Human Papilloma Virus status

Kaixian Yu [a,1], Youyi Zhang [a,1], Yang Yu [c], Chao Huang [b], Rongjie Liu [a], Tengfei Li [a], Liuqing Yang [c], Jeffrey S. Morris [a], Veerabhadran Baladandayuthapani [a], Hongtu Zhu [a,*]

[a] Department of Biostatistics, The University of Texas M.D. Anderson Cancer Center, United States
[b] Department of Biostatistics, University of North Carolina at Chapel Hill, United States
[c] Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, United States

ARTICLE INFO

Article history:
Received 27 June 2017
Revised 29 September 2017
Accepted 2 October 2017
Available online 6 November 2017

Keywords:
Radiomics
CT image
HPV status
Oropharynx cancer
Statistical method

ABSTRACT

Human Papilloma Virus (HPV) has been associated with oropharyngeal cancer prognosis. Traditionally the HPV status is tested through invasive lab test. Recently, the rapid development of statistical image analysis techniques has enabled precise quantitative analysis of medical images. The quantitative analysis of Computed Tomography (CT) provides a non-invasive way to assess HPV status for oropharynx cancer patients. We designed a statistical radiomics approach analyzing CT images to predict HPV status. Various radiomics features were extracted from CT scans, and analyzed using statistical feature selection and prediction methods. Our approach ranked the highest in the 2016 Medical Image Computing and Computer Assisted Intervention (MICCAI) grand challenge: Oropharynx Cancer (OPC) Radiomics Challenge, Human Papilloma Virus (HPV) Status Prediction. Further analysis on the most relevant radiomic features distinguishing HPV positive and negative subjects suggested that HPV positive patients usually have smaller and simpler tumors.

Introduction

Oropharyngeal cancer prognosis is often linked with Human Papilloma Virus (HPV), especially HPV type 16. HPV associated oropharynx cancer patients have been shown to have increased survival time and better tumor control with radiotherapy than non-HPV-associated ones [1–3]. Typically, HPV status is tested using immunohistochemistry for p16, a protein, or in situ hybridization for viral DNA. However, the lab testing usually requires collecting biospecimen from the patients, thus it is invasive and may impose potential risk to the patients. Therefore, seeking non-invasive yet accurate way to assess the HPV status becomes important. One possible solution is through the low dose computed tomography (CT) scans which is done routinely for screening, diagnosis, and treatment guidance. A low dose CT scan is non-invasive, and it is less likely to impose extra risks to testees. Besides non-invasiveness, the imaging technique could help the physician collect more information at diagnose which subsequently will improve the design of treatment, makes the treatment more precise for each patient.

Recently, the rapid development of radiomics has enabled more meaningful and precise quantitative analysis of medical images in various body sites. Magnetic Resonance Imaging (MRI) of brain has been extensively studied over the past two decades, many efforts have been done to quantify the relationship between radiomic features of MRI and the diagnosis of Alzheimer's disease [4–8] as well as risk of Autism [9]. CT image is another important image modality that has been widely studied to built its connection with clinical outcomes. CT scan plays a critical role in early detection and prognosis for different types of cancer [10–13]. However, only a few studies have tried to connect the CT images with HPV status quantitatively, especially in oropharynx cancers. Cantrell and colleagues [14] studied the radiomic differences in CT images between HPV+ and HPV−, but their study was not focused on predicting HPV status. Buch and colleagues [15] analyzed the difference of 42 space-invariant texture features between HPV+ and HPV− patients, again this study did not come up with a predictive model for HPV+/−. Bogowicz and colleagues [16] built a logistic regression distinguishing HPV+/− with radiomics features and achieved 0.78 AUC in validation cohort, but the number of radiomic features they used was still relative small, and not comprehensive enough. Besides that, the discovering cohort in their study was relative small.

In this article, we describe an approach that utilizes only CT image to predict HPV status for oropharynx cancer patients. Our

* Corresponding author.
E-mail address: hzhu5@mdanderson.org (H. Zhu).
[1] These authors contribute equally.

https://doi.org/10.1016/j.ctro.2017.10.001
2405-6308/

method won the first place among 9 participating teams in *the 2016 Medical Image Computing and Computer Assisted Intervention (MICCAI) grand challenge: Oropharynx Cancer (OPC) Radiomics Challenge, Human Papilloma Virus (HPV) Status Prediction*, with Area Under the Curve (AUC) 0.91549 in the held out evaluation data set. Our approach is based on statistical analysis of radiomic features of the upper chest CT. The feature selection result showed that among all the radiomic features extracted, *MeanBreadth* and *SphericalDisproportion* were the most important features that capture the most predictive radiomic information of the HPV status. We have also discovered that the results indicate that the HPV associated patients usually have smaller and simpler tumors. It is also worth mentioning that our method had a higher AUC in private leaderboard than the one in the public leaderboard, which indicates the proposed method is fairly generalizable.

Materials and methods

The analysis is based on radiomic features extracted from upper chest CT images, and prediction is assessed via statistical predictive models. The main challenges include extracting meaningful and predictive radiomic features, combining information across multiple Region Of Interests (ROI), selecting most relevant features, and constructing powerful predictive models. We have adopted existing state-of-the-art feature extraction method [17] to obtain a set of radiomic features covering various aspects of the image, for example, shape, texture, and grayscale intensities. To handle the multi-ROI situation, we computed a "consensus" ROI that can be regarded as ROI representative for each feature and for each individual subject. Statistical methods, e.g. generalized linear model (GLM), random forest, tree based model, and etc., were employed to select the relevant features and make the final prediction. The general procedure is shown in Fig. 1.

Clinical and imaging data

In this challenge, contrast-enhanced Computed Tomography (CT) scans of upper chest for 315 oropharynx cancer patients were provided as the radiomic dataset (detailed challenge setting and information of data acquisition, processing, and inclusion/exclusion, can be find in the challenge summary paper [18]). 150 randomly selected subjects with annotated HPV status were given as training cohort, with 128 HPV positive and 22 HPV negative patients. The remaining 165 subjects were held out as a validation cohort (in the actual challenge, one half of the validation subjects were scored all through the last of the challenge as the public measure of the performance, and the other half were held out as a private score which was not released until the challenge was concluded). The human papilloma virus (HPV) status defined by p16, which was tested and provided by the organizers, is the main interest of this work. Two types of ROIs, Gross Primary/Nodal Tumor volume (GTVp/GTVn) are considered, and the ROI segmentation was done manually by board certified physicians in the organizers' institution (University of Texas MD Anderson Cancer Center). In this challenge, we did not make distinguish between GTVp and GTVn per organizers' suggestion.

Pre-processing

Several randomly selected 2D slices (5–10) of the CT images of each subject were manually inspected to assure the consistence of image quality. Subjects with low quality images including overexposed and large degree of blurring were removed from the training cohort. We removed one subject (id: 88), since their CT scans were quite blurring in the main interested section, and the ROIs annotated were of volume 0.

Radiomic feature extraction

1683 radiomic features of 5 categories were extracted using IBEX [17], including Gray level cooccurrence matrix (GLCM 2D and 3D), Gray level run length matrix, intensity (and histogram), neighbor intensity difference (2D and 3D), and shape (for more details on parameters and feature types in each category please see Supplementary file 1). Quality control was done to ensure there is no missing values in extracted features. And the possible outliers, identified by Grubbs' test [19] were discarded.

One critical problem in this dataset was that one subject often had more than one ROIs (GTVp or GTVn); therefore, a proper way to "choose" a representative ROI became important and essential to the prediction accuracy. Treating each ROI as an individual subject yields scientific and practical problems due to the fact that not all ROIs of a specific subject directly reflect HPV status. In this challenge, we designed a "consensus" ROI for each subject: if there was only one ROI for some subject, then we used that ROI to represent the subject; if there were more than one ROIs for a subject, we created a virtual ROI, the ROI still had the same set of features, but the values of the features were not necessary from the same ROI but rather taking the most extreme value (extremum, in terms of magnitude) comparing to the robust median of all ROIs of all subjects in the cohort (an example is given in Fig. 2).

Statistical analysis

The general statistical analysis is outlined as follows:

1. Homogeneous testing between training and validation cohorts was done by Wilcoxon-rank-sum test [20,21] for each feature. The features with p-value <0.05 were discarded due to their inhomogeneity, which will subsequently affect the predictive model built from the training cohort.
2. Preliminary feature screening was done by Kolmogorov–Smirnov (KS) test [22,23]. For each feature, calculating the KS statistic between HPV+ and HPV− subjects in the training cohort, keep only the features with p-value <0.05 to select only the features that were able to distinguish the two groups.
3. Further feature screening through correlation with HPV status. A biserial correlation between HPV status and each radiomic feature was calculated, and only the features with biserial absolute correlation >0.3 were kept to achieve clinical relevance.
4. Ranking remaining features by their marginal Area Under Curve (AUC) obtained by 10-time random split. The marginal AUC was accessed by building model with only one feature. The random split was done by randomly sampling 50% of the training cohort to train the model, then evaluating on the held out 50% data. We have tested various statistical models, including generalized linear model (GLM) [24], pdfCluster [25], predictive tree model [26], random forest [27], Support Vector Machine (SVM) [28], and etc. Only the top 10 features with highest marginal AUC were kept in this step.
5. The final features in the model was selected by forward selection, where we add features one by one according to their ranks of marginal AUC from high to low, until the model AUC stopped increasing. The model AUC was accessed by 10-time random split as well.

Selecting appropriate statistical model (tree, GLM, SVM,...) was done by submitting to the public leaderboard of this challenge, as well tuning some parameters. e.g. number of trees in random forest.

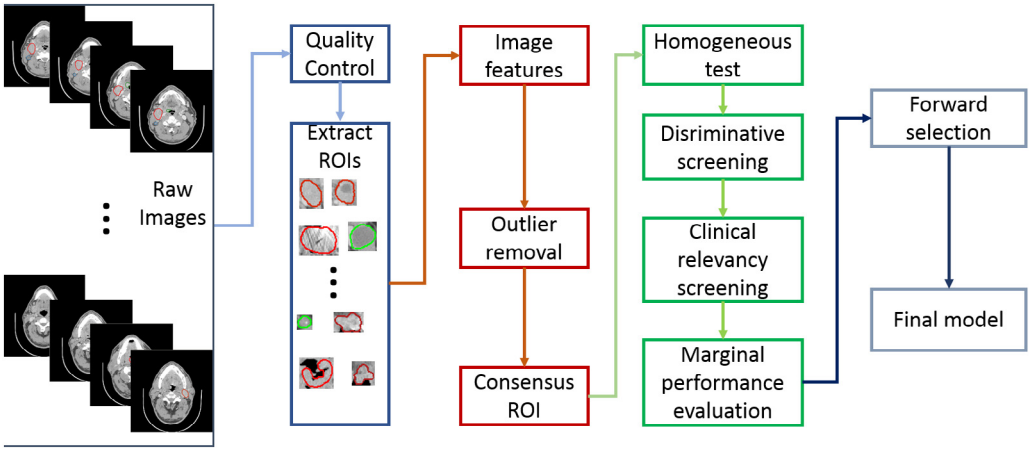
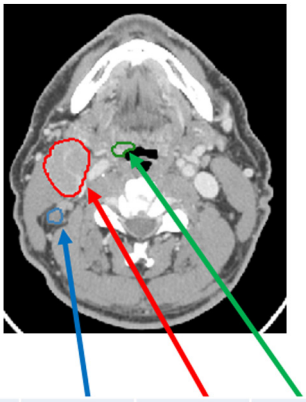


Fig. 1. The overall procedure of the wining method.



Features*	ROI1	ROI2	ROI3	Cohort Median	Consensus
Contrast	382.4232	432.0918	835.9087	570.3639	835.9087
TextureStrength	0.363721	0.100572	1.673612	0.891689	1.673612
Compactness1	0.247417	0.830998	0.184217	0.342112	0.830998
Compactness2	0.354137	0.541193	0.775638	0.45023	0.775638
Convex	0.892999	0.963741	0.970771	0.951638	0.892999
ConvexHullVolume	2.610085	17.69994	1.012324	4.176374	17.69994
ConvexHullVolume3D	3.264901	20.24863	1.025358	4.744803	20.24863
Mass	2.62334	18.75935	1.096457	4.410034	18.75935
Max3DDiameter	2.215496	4.675022	1.497448	3.018924	4.675022
MeanBreadth	2.147786	3.891796	1.26263	2.560937	3.891796
...	...	...	...	...	...

*Some features were calculated based on the 3D image, here only one slice is shown.

Fig. 2. The creation of consensus ROI out of a subject with 3 ROIs.

Results

Eventually, we selected the highest performed model, logistic regression with two features *MeanBreadth* and *SphericalDisproportion* (Table 1). The performance metric used in this challenge was Area Under Curve (AUC), the model obtained mean AUC of 0.753 (±0.075) from 10-random-split on training cohort (Fig. 3), 0.86667 on public leaderboard, and 0.91549 on the private one. We are the only team having higher private leaderboard score than the public one, which is a good indicator of model generalizability. Due to the challenge setting that truth were not released for public and private leaderboard cohorts, we are not able to provide ROC curve for either leaderboard cohorts.

Table 1
Final logistic regression.

Variable	Estimated odds ratio	95% CI	*p*-value
MeanBreadth	0.926	[0.895, 0.958]	<0.0001
SphericalDisproportion	2.045	[1.833, 2.280]	<0.0001

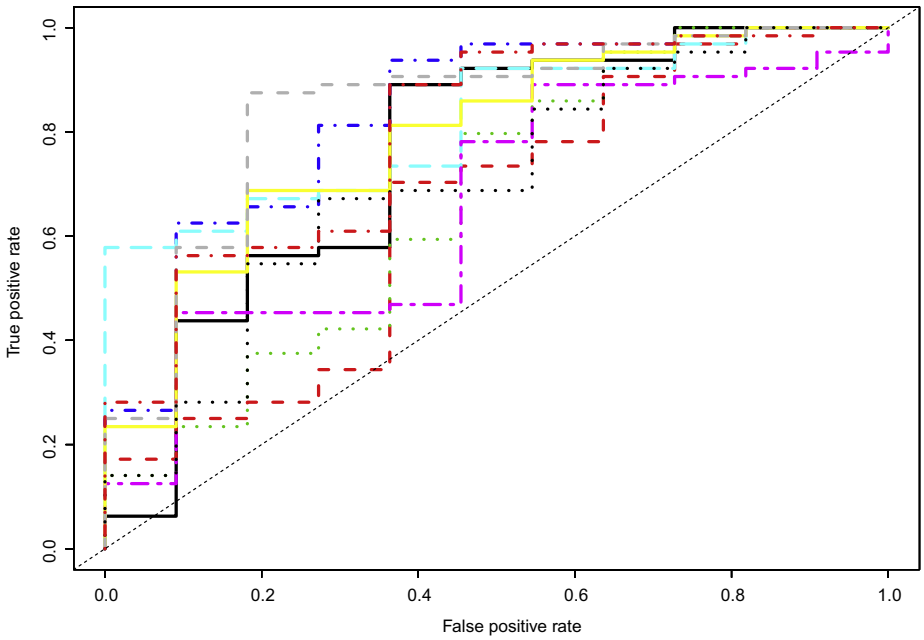


Fig. 3. The testing ROC curves of each fold of the 10-random-split of the training cohort.

The *MeanBreadth* measures the mean "width" of the ROI; therefore, it is closely related to the size of the tumor. We observed that HPV-pos subjects tend to have smaller meanBreadth (Fig. 4A), which indicates the tumor size is smaller in the HPV induced subjects.

On the other hand, *SphericalDisproportion* measures the ratio of the surface area of the image ROI to the surface area of a sphere with the same volume as the image ROI. This feature describes how complicated the shape of the tumor is, since simpler shape usually have smaller surface area comparing to more complicated shape, when the volumes are the same. The HPV-positive subjects also have smaller *SphericalDisproportion* (Fig. 4B), which could imply that the shape of the tumor is less complex for subjects carrying HPV than the one for subjects not carrying HPV.

Discussion

In this article, we reported our winning strategy in the 2016 Medical Image Computing and Computer Assisted Intervention (MICCAI) grand challenge: Oropharynx Cancer (OPC) Radiomics Challenge, Human Papilloma Virus (HPV) Status Prediction. The goal of the challenge is to predict HPV-16 status from annotated CT images. Our approach involves image quality checking, feature extraction, ROI reconstruction and variable selection. All through the process, we have tried various statistical models including pdfCluster, random forest, decision tree, SVM (linear and non-linear kernels) and etc. In the winning submission, we used the generalized linear model (GLM) since it had the best public leaderboard score.

Besides the statistical models, we have tried deep learning as well. The network we used was GoogleNet, and the input was 2D center slice of the ROI. Although the results turned out to be not as satisfied as the statistical models, it is worth mentioning that even with only 149 subjects (actually only 75 subjects since the random split takes only half of the data as training cohort), deep learning was able to achieve AUC around 0.744 on a random split test (0.753 for the winning algorithm on the same set). The difference in AUC is marginal, and we expect that given more training subjects, deep learning will behave as well as if no better than the statistical models.

Several clinical statements were also provided, but we did not use these information in our model mainly due to two reasons: firstly, our intuition was to assess the performance of using images solely, and evaluate how precise image itself can predict the HPV status; another one is including clinical information will bring in extra uncertainties into the model. For example, TNM grade requires further testing, and the grade system is discrete while tumors develop in a continuous fashion; therefore, a cutoff has to be chosen to classify the tumor into the grade system. This mandatory discretization may introduce bias and uncertainty. On the other hand, some self-reported items, e.g. smoking and sexual frequency may not be accurate. Therefore, if the clinical parameters will help improve the prediction is still debatable. There was a study [29] showing that the clinical parameters themselves were able to achieve prediction AUC of 0.84 (lower than what our model can do). They have also showed that their radiomic features did not provide improvement over the clinical only model. However, they have had only a few vaguely defined imaging features, while in our study much more comprehensive radiomic features were extracted. Unfortunately, since the challenge was concluded, we no longer have access to the leaderboard, so we are not able to provide feedbacks at this time. But this issue, whether or not adding clinical parameters improve the overall prediction accuracy, worth further study.

One uncommon fact is that our model has a higher AUC on the private cohort than the public cohort. On one hand we believe it is an indication of good generalizability of our model, while it could also be possible that the validation/test cohort is relatively small. Due to the setting of the challenge that we did not receive the truth of either public or private cohort, we are not able to give more information on this rare situation.

From the results, we found that the subjects of HPV-pos usually have smaller and geometrically simpler tumor, hence it is more manageable. These findings may partially explain the current literature results that HPV-pos subjects have overall survival advantage [1–3].

In this study, we explored the correlation between radiomic features and one clinical outcome, HPV16 status. As an exploration study, we think it provides its value by showing the high predictive ability. Beyond HPV16 status, we believe there are more

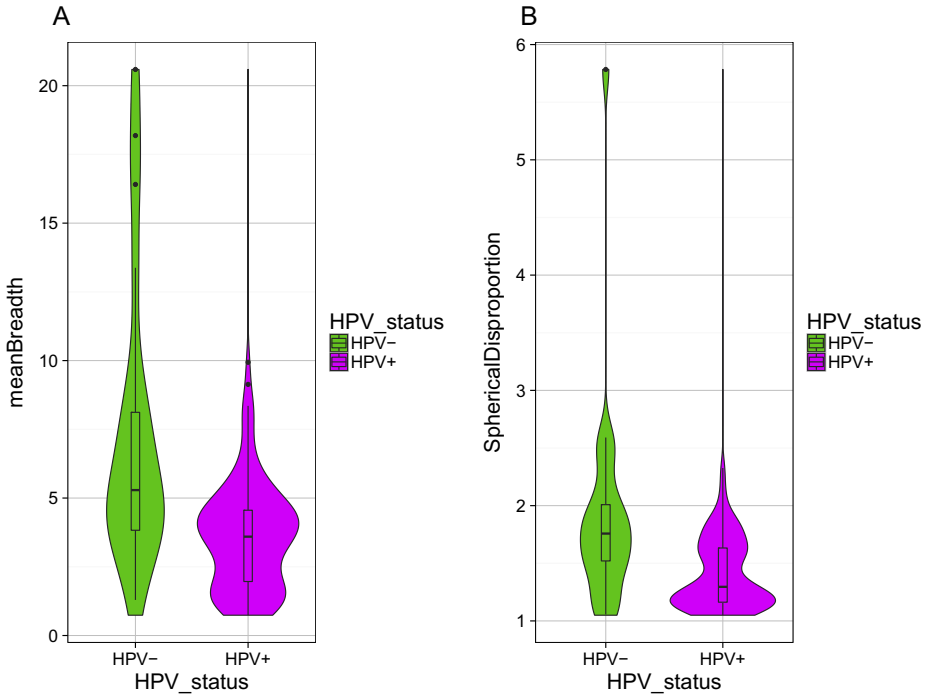


Fig. 4. Features comparison between HPV+ and HPV− subjects. A. HPV+ patients have a relatively smaller *meanBreadth* comparing to HPV− patients. B. Similarly, HPV− subjects have larger *SphericalDisproportion* than the HPV+ subjects.

connections between the underlying biology and imaging, for example quantifying tumor heterogeneity through imaging. In this scenario, clonal/subclonal composition can be identified through imaging, without getting the actual biopsy. This will help physician make better treatment decision, and potentially increase the survival chance of patients.

In conclusion, we have designed a statistical framework to analyze CT images to predict HPV status, and achieved the first place in the 2016 MICCAI grand challenge: Oropharynx Cancer (OPC) Radiomics Challenge, Human Papilloma Virus (HPV) Status Prediction.

Acknowledgments

We would like to thank the editor and reviewers for their constructive comments to help us improve this manuscript. HZ's research was partially supported by NIH Grant MH086633, NSF Grants SES-1357666 and DMS-1407655, and a Grant from Cancer Prevention Research Institute of Texas. VB's research was partially supported by NIH Grants R01-CA194391 and R01160736, NSF Grant 1463233 and Cancer Center Support Grant (CCSG) from NIH/NCI (P30-CA016672). JSM's efforts were supported by NCI Grants CA-178744, CA-160736, and CA-016672, and NSF-Bio/BBSRC Grant 1550088. We would like to thank the organizers: Clifton D Fuller, MD, PhD (Team Lead), Hesham ElHalawani, MD, MSc, Abdallah Mohamed, MD, MSc, Aubrey White, BS, James Zafereo, BS, Andrew Wong, BS, Joel Berends, BS, Shady Abohashem, MD, Bowman Williams, BS, Jeremy Aymard, BS, Aasheesh Kanwar, BS and Subha Perni, BS (MD Anderson Cancer Center); MGH: Jayashree Kalpathy-Cramer,PhD; NCI: Keyvan Farahani, PhD, John Freymann, PhD for making this challenge possible.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ctro.2017.10.001.

References

[1] de Jong MC, Pramana J, Knegjens JL, Balm AJ, van den Brekel MW, Hauptmann M, et al. Hpv and high-risk gene expression profiles predict response to chemoradiotherapy in head and neck cancer, independent of clinical factors. Radiother Oncol 2010;95(3):365–70. https://doi.org/10.1016/j.radonc.2010.02.001.

[2] Ang KK, Harris J, Wheeler R, Weber R, Rosenthal DI, Nguyen-Tan PF, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. N Engl J Med 2010;363(1):24–35. https://doi.org/10.1056/NEJMoa0912217.

[3] Fakhry C, Westra WH, Li S, Cmelak A, Ridge JA, Pinto H, et al. Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial. J Natl Cancer Inst 2008;100(4):261–9. https://doi.org/10.1093/jnci/djn011.

[4] Kong D, Giovanello KS, Wang Y, Lin W, Lee E, Fan Y, et al. Predicting alzheimers disease using combined imaging-whole genome snp data. J Alzheimer's Dis: JAD 2015;46(3):695–702. https://doi.org/10.3233/JAD-150164.

[5] Jack Jr CR, Petersen RC, Xu YC, O'Brien PC, Smith GE, Ivnik RJ, et al. Prediction of ad with mri-based hippocampal volume in mild cognitive impairment. Neurology 1999;52(7):1397–403.

[6] Fennema-Notestine C, Hagler Jr DJ, McEvoy LK, Fleisher AS, Wu EH, Karow DS, Dale AM. Structural mri biomarkers for preclinical and mild alzheimer's disease. Hum Brain Mapping 2009;30(10):3238–53. https://doi.org/10.1002/hbm.20744.

[7] Davatzikos C, Bhatt P, Shaw LM, Batmanghelich KN, Trojanowski JQ. Prediction of mci to ad conversion, via mri, csf biomarkers, and pattern classification. Neurobiol Aging 2011;32(12). https://doi.org/10.1016/j.neurobiolaging.2010.05.023. 2322.e19–27.

[8] Cuingnet R, Gerardin E, Tessieras J, Auzias G, Lehericy S, Habert MO, et al. Automatic classification of patients with alzheimer's disease from structural mri: a comparison of ten methods using the adni database. Neuroimage 2011;56(2):766–81. https://doi.org/10.1016/j.neuroimage.2010.06.013.

[9] Hazlett HC, Gu H, Munsell BC, Kim SH, Styner M, Wolff JJ, et al. Early brain development in infants at high risk for autism spectrum disorder. Nature 2017;542(7641):348–51.

[10] Caroline C. Lung cancer screening with low dose ct. Radiol. Clinics North Am. 2014;52(1):27–46. https://doi.org/10.1016/j.rcl.2013.08.006.

[11] Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 2011;365(5):395–409. https://doi.org/10.1056/NEJMoa1102873.

[12] Lindfors KK, Boone JM, Nelson TR, Yang K, Kwan ALC, Miller DF. Dedicated breast ct: initial clinical experience. Radiology 2008;246(3):725–33. https://doi.org/10.1148/radiol.2463070410.

[13] Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014;5:4006.

[14] Cantrell SC, Peck BW, Li G, Wei Q, Sturgis EM, Ginsberg LE. Differences in imaging characteristics of hpv-positive and hpv-negative oropharyngeal cancers: a blinded matched-pair analysis. AJNR. Am J Neuroradiol 2013;34 (10):2005–9. https://doi.org/10.3174/ajnr.A3524.
[15] Buch K, Fujita A, Li B, Kawashima Y, Qureshi M, Sakai O, Using texture analysis to determine human papillomavirus status of oropharyngeal squamous cell carcinomas on ct. Am J Neuroradiol36(7); 2015, pp. 1343–1348. arXiv: http://www.ajnr.org/content/36/7/1343.full.pdf.https://doi.org/10.3174/ajnr.A4285. URLhttp://www.ajnr.org/content/36/7/1343.
[16] Bogowicz M, Riesterer O, Ikenberg K, Stieb S, Moch H, Studer G. et al., Computed tomography radiomics predicts hpv status and local tumor control after definitive radiochemotherapy in head and neck squamous cell carcinoma. Int J Radiat Oncol Biol Phys.https://doi.org/10.1016/j.ijrobp.2017.06.002. URLhttp://www.sciencedirect.com/science/article/pii/S0360301617310118.
[17] Zhang L, Fried DV, Fave XJ, Hunter LA, Yang J, Court LE. Ibex: an open infrastructure software platform to facilitate collaborative work in radiomics. Med Phys 2015;42(3):1341–53.
[18] Matched computed tomography segmentation and demographic data for oropharyngeal cancer radiomics challenges 4 (2017) 170077.https://doi.org/10.1038/sdata.2017.77.https://www.nature.com/articles/sdata201777#supplementary-information.
[19] Grubbs FE, Procedures for detecting outlying observations in samples, Technometrics 11 (1), c6758 Times Cited:1461 Cited References Count:15.
[20] Wilcoxon F. Individual comparisons by ranking methods. Biometrics Bull 1945;1(6):80–3. https://doi.org/10.2307/3001968.
[21] Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. Ann Math Stat 1947:50–60. https://doi.org/10.1214/aoms/1177730491.
[22] Kolmogorov AN. Sulla determinazione empirica di una legge di distribuzione. Giornale dell'Istituto Italiano degli Attuari 1933;4:83–91. doi:citeulike-article-id:1213059.
[23] Smirnov N. Table for estimating the goodness of fit of empirical distributions. Ann Math Stat 1948:279–81. https://doi.org/10.1214/aoms/1177730256.
[24] Nelder JA, Wedderburn RWM. Generalized linear models. J R Stat Soc Ser A (General) 1972;135(3):370–84. https://doi.org/10.2307/2344614.
[25] Menardi G, Azzalini A. An advancement in clustering via nonparametric density estimation. Stat Comput 2014;24(5):753–67. https://doi.org/10.1007/s11222-013-9400-x.
[26] Quinlan JR. Learning efficient classification procedures and their application to chess end games. Berlin Heidelberg: Springer; 1983. p. 463–82.
[27] Breiman L. Random forests. Mach Learn 2001;45(1):5–32. https://doi.org/10.1023/A:1010933404324.
[28] Boser BE, Guyon IM, Vapnik VN, A training algorithm for optimal margin classiffers; 1992.https://doi.org/10.1145/130385.130401.
[29] Chan MW, Yu E, Bartlett E, O'Sullivan B, Su J, Waldron J, et al. Morphologic and topographic radiologic features of human papillomavirus-related and unrelated oropharyngeal carcinoma. Head Neck 2017;39(8):1524–34. https://doi.org/10.1002/hed.24764.