The Prognostic Value of Foundational Radiomics and Clinical Features in Locally Advanced Non-Small Cell Lung Cancer in NRG Oncology Trial RTOG 0617

Taman Upadhaya, Elizabeth M. McKenzie, Indrin J. Chetty and Katelyn M. Atkins

Department of Radiation Oncology, Cedars-Sinai Medical Center, Los Angeles, USA

Abstract

This study investigated the feasibility of the foundational radiomics technique on the downstream task of predicting overall survival outcomes in patients with locally advanced non-small cell lung cancer (NSCLC). In an analysis of 449 patients from NRG Oncology/Radiation Therapy Oncology Group (ROTG) 0617, univariate cox regression identified 241 foundational radiomic and 19 clinical features as significant (p-value≤0.05) independent predictors of survival. Subsequently, the performance of popular predictive models, such as Least Absolute Shrinkage and Selection Operator (LASSO) in conjunction with a generalized linear model (GLM), Random Forest (RF), and Support Vector Machine (SVM) were evaluated using 5-repeat 10-fold cross-validation (n=278) and in an unseen independent test set (n=140), yielding an AUC of 0.67 (95% confidence interval [CI], 0.64-0.70) and 0.64 (95% CI, 0.54-0.73) based on foundational radiomics model solely employing SVM and 0.66 (95% CI, 0.73-0.68) and 0.67 (95% CI, 0.58-0.76) based on the clinical features model solely employing LASSO/GLM, respectively. When the foundational radiomics features were combined the clinical features, the AUC using a RF model increased to 0.74 (95% CI, 0.72-0.77) for the validation datasets and 0.67 (95% CI, 0.58-0.76) for the unseen independent dataset. This suggests that radiomics signatures provide complimentary information to clinical features for prediction of survival outcomes in patients with locally advanced NSCLC.

1. Introduction

A foundational model is an artificial intelligence (AI) model that uses a self-supervised learning technique to build the deep learning model on a vast amount of unlabeled data that can be applied to a wide range of downstream use cases. Recently, foundational models have demonstrated significantly enhanced performance in language and vision, driving successes in applications of ChatGPT, BERT, and CLIP in natural language processing (NLP), and SimCLR, YOLO, DINO, SAM and DALLE in computer vision [1, 2]. Furthermore, leveraging a foundational model for downstream tasks helps address the challenge of both limited and labeled data for modeling. This is particularly true in the domain of medical image analysis as colleting a substantial amount of data for modeling can be challenging. Thus, a foundational model capable of learning from diverse datasets through self-supervised learning, independent of specific outcomes, and capable of predicting various downstream tasks is highly desirable. To date, the only known foundational model study for 3D medical imaging was published by Pai *et el.,* [3] which was trained using a comprehensive dataset of 11,467 radiographic cancers on computed tomography (CT) imaging for various clinical relevance tasks. While there are many studies related to hand crafted radiomics features (intensity, texture, morphological) and deep radiomics (based on deep learning) the research for foundation radiomics (based on a foundational model) for prediction of oncologic outcomes is largely unexplored [4, 5].

In this work we exploit a pretrained foundational model from Pai *et el.,* for the downstream task of predicting 2-year overall survival in patients with locally advanced NSCLC. Our approach utilizes the foundational model as a feature extractor with various minor and major modification. In addition, we evaluated the complementary value of foundational radiomics features with clinical features using popular machine learning (ML) classifier namely generalized linear model with LASSO penalty, Random Forest (RF) and Support Vector Machine (SVM) for their performance.

2. Materials and Methods

2.1. Clinical cohort data

The NRG Oncology/Radiation Therapy Oncology Group (ROTG) 0617 data set from the National Clinical Trials Network (NCTN) was utilized. [6]. All relevant CT images, RT plans and clinical features such as Zubrod performance status, tumor histology, age, race, gender, smoking history, pneumonitis, esophagitis, chemotherapy, cetuximab, and RT dose groups (60 vs. 74 Gy) were utilized as per the NCTN dataset [7]. Additionally, dose statistics from heart, left anterior descending coronary artery [8], planning tumor volume (PTV), clinical tumor volume CTV), and ipsilateral lung structures were included in the analysis. Survival time and vital status were used to estimate 2-year overall survival (OS).

2.2. Foundational radiomics extraction

We employed a pretrained foundational model developed by Pai *et el.,* which was pre-trained on 11,467 diverse annotated cancers identified on computed tomography (CT) imaging. This foundational model is the modified version of the original SimCLR framework to effectively represent the nature of medical images [9]. It adopts ResNet50 architecture as encoder with the contrastive loss function in a self-supervised manner to learn to identify positive pairs of 3D patches surrounding the lesion and negative pairs, randomly sampled 3D patches from the rest of the scan [3, 10]. Once the model is trained, it can be used on various

downstream tasks, such as demonstrated in the original study. Furthermore, for our specific task, even though we used the same pretrained model, we modified the image transformation step by removing the seed-based crop to lesion-based crop to accommodate cancers of various sizes and multiple lesions within the region of interest. We then used these regions of interest from the CT images to extract 4098 features for our downstream task of predicting 2-year OS, which was termed features foundational radiomics.

2.3 Feature selection and machine learning

Univariate cox regression and correlation were applied to identify significant and independent predictors of 2-year OS. Subsequently, popular predictive models, such as LASSO, RF, and SVM were trained. Pre-processing for removing zero-variance features and centering (mean of zero and standard deviation of one) was applied before feeding features to the algorithms. Class imbalance was handled using the Synthetic Minority Over-sampling Technique (SMOTE) and tuning of the parameters for SVM, RF and GLM/LASSO were done using 5-repeat 10-fold cross-validation. All independent predictors were rank based on the importance score from LASSO, RF and SVM and then stepwise forward feature selection based on the previous ranking is used to select the subset of features minimizing the validation error. The final model for testing was chosen based on the highest area under the curve (AUC). For technical details, we direct readers to our prior publications [11, 12]. Three distinct experimental setups were carried out: (i) in the first approach, all RTOG 0617 patients (n=449) were included (for both high and low dose treatment arms), and after preprocessing split into training-testing cohorts (n=278 training, n=140 testing); (ii) the second approach was limited to patients from high dose arm (n=184 training, n=46 testing); and (iii) the third approach involved was limited to patients from low dose arm (n=150 patients training, n=38 testing).

3. Results

The univariate Cox analysis identified 822 foundational radiomics features (hazard ratios ranging from 0.14 to 3.35, and corresponding p-values ≤0.05) and 26 clinical features (hazard ratios ranging from 0.56 to 6.26 and p-values ≤0.05), as significant. Among them, 241 foundational radiomics features, along with 19 clinical features, were identified as independent predictors of survival.

Figure 1 illustrates the comprehensive performance results derived from our experiments. In the initial set of experiments, which incorporated all patients regardless of radiation dose, the model leveraging solely 12 foundational radiomics features achieved an AUC of 0.67 (95% confidence interval [CI], 0.64-0.70) in cross-validation and 0.64 (95% CI, 0.54-0.73) in unseen independent testing

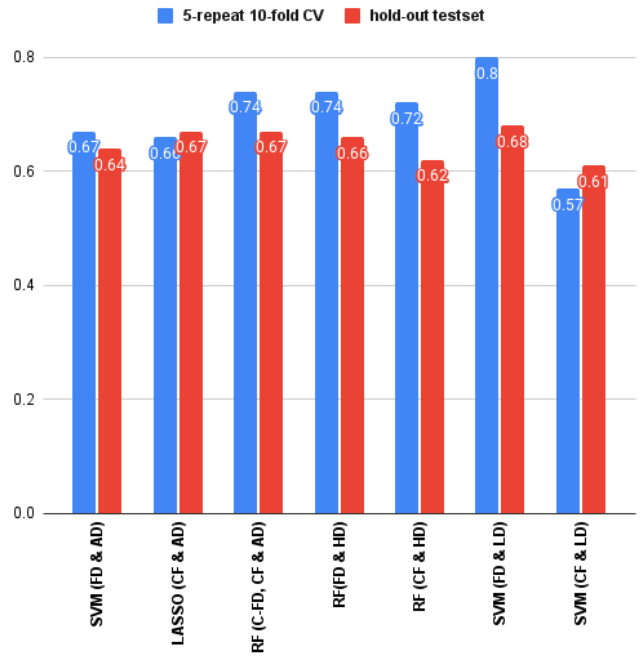

Figure 1. Performance of foundational radiomics feature (FD) based model, clinical feature (CF) based model and their combination (C-FD and CF) using all patients irrespective of dose (AD), as well as separately within high dose (HD) arm and low dose (LD) arm as assessed by 5-repeat 10-fold cross validation and in unseen independent test set.

using the SVM algorithm. In contrast, the model relying solely on 6 clinical features achieved an AUC of 0.66 (95% CI, 0.73-0.68) in cross-validation and 0.67 (95% CI, 0.58-0.76) in unseen independent testing using the LASSO algorithm. The combination of 7 foundational radiomics features and 4 clinical features (grade 3 esophagitis, PTV volume, ipsilateral lung mean dose, v5 heart and v5 lung percentage) produced the highest AUC of 0.74 (12 features) (95% CI, 0.72-0.77) in cross-validation and 0.67 (95% CI, 0.58-0.76) in unseen independent testing RF algorithm.

Similarly, in the subsequent set of experiments involving patients from the high dose arm only, the model based solely on 5 foundational radiomics features achieved the highest AUC of 0.74 (95% CI, 0.70-0.767) in cross-validation and 0.66 (95% CI, 0.50-0.83) in unseen independent testing using the RF algorithm. Conversely, the model relying solely on 6 clinical features (age, PTV log volume, chemotherapy, Left Anterior Descending Coronary Artery (LAD) mean dose, v5 lung and V95 PTV percentage) achieved an AUC of 0.72 (95% CI, 0.69-0.75) in cross-validation and 0.62 (95% CI, 0.45-0.78) in unseen independent testing, once again using the RF algorithm. However, the integration of foundational radiomics and clinical features did not enhance the AUC performance.

In the final experiment which incorporated patients from the low dose arm only, the model relying solely on 12 foundational radiomics achieved the highest AUC of 0.80 (95% CI, 0.77-0.83) in cross-validation and 0.68 (95% CI, 0.50-0.85) in unseen independent testing using the SVM algorithm. Meanwhile, the model relying solely on 9 clinical features (PTV log volume, ipsilateral lung mean dose, grade5 toxicity, LAD mean dose, heart mean dose, heart median dose, V5 and V20 lung percentage, V60 esophagus) achieved the lowest performance, with an AUC of 0.57 (95% CI, 0.53-0.61) in cross-validation and 0.61 (95% CI, 0.42-0.79) in unseen independent testing, once again using the SVM algorithm. Similar to above, the combination of foundational radiomics and clinical features did not enhance the AUC performance.

4. Discussion

In this study, we demonstrated that prediction models for OS can be built using foundational radiomics with ML techniques. In our first experiments involving all patients irrespective of dose received, the model incorporating both foundational radiomics and clinical features outperformed the model replying solely on foundational radiomics features or clinical features. Overall, all methods led to models with modest AUC (around or above 0.66)**.** As we shifted our approach from dose-agnostic to those receiving high vs low dose treatments separately, we observed noticeable variations in model performance, worth of further study. For the experiment involving only patients from the high-dose arm, the foundational radiomics-based model solely achieved a performance of 0.74 in cross-validation and 0.66 in the unseen test set, outperforming the clinical feature-based model. Similarly in the experiment involving only patients from the low dose arm, the foundational radiomics-based model reached a performance of 0.80 in cross-validation and 0.68 in unseen independent test set, again outperforming the clinical features-based model. In both analyses by dose arm, the combination of foundational radiomics and clinical features did not improve the model performance.

To our best knowledge, this study is the first to investigate the prognostic and complementary value of foundational radiomics with clinical features using various popular machine learning algorithms. To date, only one study has been reported using foundational radiomics for prognostication by Pai et al., however they only included one linear classifier. Our study achieved similar performance to the previous study using foundational radiomics, though our approach uniquely explored the additional benefits of combining these foundational radiomics features with clinical features.

Our study has notable limitations. In the model development, we did not perform fine-tuning on our data set that could potentially improve the performance. However, it is important to note as these foundational models get bigger- hosting, building, and training a foundational is an incredibly expensive endeavor—often costing millions of dollars. Therefore, investigating downstream tasks of models such as these would be crucial. Our analysis also did not include handcrafted radiomics features. We are actively working on expanding our work to fine tune the foundational model, include handcrafted radiomics features and additional anatomical structures such as the heart and lung, which could provide further insights into tissue characteristics, as our ML framework is easily amenable to incorporation of additional data [13].

5. Conclusion

In conclusion, a self-supervised foundational model based on radiomics signatures showed promising prediction of 2-year OS for patients with locally advanced NSCLC from RTOG 0617. The combination of foundational radiomics with conventional clinical features provided even better prediction, suggesting the complimentary value of foundational radiomics in outcomes modeling.

References

[1] Wang, H., Li, J., Wu, H., Hovy, E. and Sun, Y., 2022. Pre-trained language models and their applications. Engineering. DOI: https://doi.org/10.1016/j.eng.2022.04.024

[2] Zhang, J., Huang, J., Jin, S. and Lu, S., 2023. Vision-language models for vision tasks: A survey. arXiv preprint arXiv:2304.00685. DOI: https://doi.org/10.48550/arXiv.2304.00685

[3] Pai, S., Bontempi, D., Prudente, V., Hadzic, I., Sokač, M., Chaunzwa, T.L., Bernatz, S., Hosny, A., Mak, R.H., Birkbak, N.J. and Aerts, H.J., 2023. Foundation models for quantitative biomarker discovery in cancer imaging. medRxiv.

[4] Aerts, H.J., Velazquez, E.R., Leijenaar, R.T., Parmar, C., Grossmann, P., Carvalho, S., Bussink, J., Monshouwer, R., Haibe-Kains, B., Rietveld, D. and Hoebers, F., 2014. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nature communications, 5(1), p.4006.

[5] Afshar, P., Mohammadi, A., Plataniotis, K.N., Oikonomou, A. and Benali, H., 2019. From handcrafted to deep-learning-based cancer radiomics: challenges and opportunities. IEEE Signal Processing Magazine, 36(4), pp.132-160.

[6] Bradley, J.D., Hu, C., Komaki, R.R., Masters, G.A., Blumenschein, G.R., Schild, S.E., Bogart, J.A., Forster, K.M., Magliocco, A.M., Kavadi, V.S. and Narayan, S., 2020. Long-term results of NRG oncology RTOG 0617: standard-versus high-dose chemoradiotherapy with or without cetuximab for unresectable stage III non–small-cell lung cancer. Journal of Clinical Oncology, 38(7), p.706.

[7] Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M. and Tarbox, L., 2013. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. Journal of digital imaging, 26, pp.1045-1057.

[8] McKenzie, E., Zakariaee, R., Zhang, S., Guthier, C.V., Hakimian, B., Mirhadi, A.J., Kamrava, M., Padda, S., Lewis, J.H., Nikolova, A. and Mak, R.H., 2022. Left Anterior Descending Coronary Artery Radiation Dose Association With All-Cause Mortality in NRG Oncology Trial RTOG 0617. International Journal of Radiation Oncology, Biology, Physics, 114(3), p.S58.

[9] Chen, T., Kornblith, S., Norouzi, M. and Hinton, G., 2020, November. A simple framework for contrastive learning of visual representations. In International conference on machine learning (pp. 1597-1607). PMLR.

[10] Uemura, T., Näppi, J.J., Hironaka, T., Kim, H. and Yoshida, H., 2020, March. Comparative performance of 3D-DenseNet, 3D-ResNet, and 3D-VGG models in polyp detection for CT colonography. In Medical Imaging 2020: Computer-Aided Diagnosis (Vol. 11314, pp. 736-741). SPIE.

[11] Morin, O., Vallières, M., Braunstein, S., Ginart, J.B., Upadhaya, T., Woodruff, H.C., Zwanenburg, A., Chatterjee, A., Villanueva-Meyer, J.E., Valdes, G. and Chen, W., 2021. An artificial intelligence framework integrating longitudinal electronic health records with real-world data enables continuous pan-cancer prognostication. Nature Cancer, 2(7), pp.709-722.

[12] Upadhaya, T., Vallières, M., Chatterjee, A., Lucia, F., Bonaffini, P.A., Masson, I., Mervoyer, A., Reinhold, C., Schick, U., Seuntjens, J. and Le Rest, C.C., 2018. Comparison of radiomics models built through machine learning in a multicentric context with independent testing: identical data, similar algorithms, different methodologies. IEEE Transactions on Radiation and Plasma Medical Sciences, 3(2), pp.192-200.

[13] Uemura, T., Näppi, J.J., Hironaka, T., Kim, H. and Yoshida, H., 2020, McKenzie, E., Zhang, S., Zakariaee, R., Guthier, C.V., Hakimian, B., Mirhadi, A., Kamrava, M., Padda, S.K., Lewis, J.H., Nikolova, A. and Mak, R.H., 2023. Left anterior descending coronary artery radiation dose association with all-cause mortality in NRG oncology trial RTOG 0617. International Journal of Radiation Oncology* Biology* Physics, 115(5), pp.1138-1143.