



OPEN Foundation model based prediction of lung cancer survival using temporal changes in dual time point CT scans

Jessica Petrochuk¹, Suraj Pai^{1,2,3}, John He^{1,2}, Fridolin Haugg^{1,2}, Yiwen Xu^{4,5}, David Christiani^{1,2}, Raymond Mak^{1,2}✉ & Hugo Aerts^{1,2,3}

Lung cancer remains a significant cause of mortality, with non-small cell lung cancer (NSCLC) representing most cases. Currently, clinical data based models fall short in predicting survival while more advanced deep learning based image models require vast amounts of data and are often limited to predictions based on single time points. This study uses dual time point CT scans and features derived from a foundation model to predict survival. A dataset containing 102 NSCLC patients treated with radiation therapy was used, with each patient having both pre-treatment and post-treatment CT scans. A foundation model applied to the scans generated high-dimensional feature vectors and these vectors were then further summarized. Statistical analyses, including random forest and gradient boosted survival models, were then used to predict survival. The results demonstrated that temporal changes in feature vectors, specifically the Euclidean distance and element-wise subtracted feature vectors, can offer improved prediction of survival over single-time point features and clinical data.

According to the American Cancer Society, one in five cancer deaths every year in the United States occurs from lung cancer, making it one of the leading causes of cancer-related deaths and the second most common form of cancer for both men and women¹. Within these lung cancer cases approximately 84% are non-small cell lung cancer (NSCLC) with a 5-year survival rate of 26.4%^{1,2}. Radiation therapy is commonly used to treat locally advanced NSCLC in conjunction with chemotherapy, immunotherapy and/or surgery. Following treatment, response to radiation therapy is most often quantified through CT-based measurements such as those outlined in the response evaluation criteria in solid tumors (RECIST) specifications³. And yet, while RECIST criteria are the current gold standard, studies have indicated that they tend to lack predictive power^{4–6}. Furthermore, RECIST criteria increasingly become difficult to use when patients are treated with radiation therapy due to various radiation-induced lung changes such as post-treatment fibrosis⁷. Moreover, extracting this information requires manual measurements of tumors from medical imaging which is not only time-consuming but is also prone to human error and subjectivity.

Advances in artificial intelligence, specifically in the domain of computer vision, now present an alternative for extracting predictive features from imaging data. More specifically, features that can capture a higher degree of complexity than those extracted by hand. The largest drawback of this approach is that building and training reliable computer vision models requires large-scale, heavily curated datasets which can be difficult to obtain, especially within the medical imaging domain. One solution for this is medical imaging foundation models, which are large, general-purpose models pre-trained on large-scale datasets that can serve as a starting point for downstream tasks. Several medical imaging foundation models have already demonstrated their ability to apply pre-trained knowledge to tasks reliant on smaller datasets^{8–10}. Most notably, Pai et al. introduced a foundation model trained on over 11,000 radiographic lesions and validated its efficacy on a variety of clinical use-cases including predicting malignancy and prognostics¹⁰.

And yet, while these foundation models have already shown their capabilities of extracting meaningful features from medical imaging data, it is important to note that, thus far, most have been trained and tested on images from a single time point. But, within the clinic, CT scans are often a routine part of care during both

¹Department of Radiation Oncology, Dana-Farber Cancer Institute/Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. ²Artificial Intelligence in Medicine (AIM) Program, Mass General Brigham, Harvard Medical School, Harvard Institutes of Medicine, Boston, USA. ³Radiology and Nuclear Medicine, CARIM and GROW, Maastricht University, Maastricht, the Netherlands. ⁴BC Cancer, Kelowna, British Columbia, Canada. ⁵University of British Columbia, Kelowna, British Columbia, Canada. ✉email: rmak@mgb.org

initial staging and follow-up when assessing treatment response, thereby providing additional imaging data that is often not leveraged.

Combining the advantages of foundation models with the valuable information coming from temporal changes across CT scans, this study applies the state-of-the-art foundation model introduced by Pai et al. to predict time-to-event for patients with NSCLC using dual time-point CT scans before a patient received radiation therapy and after¹⁰. In this work we report the first to-our-knowledge application of multi-time-point CT imaging data in combination with a foundation model for outcome prediction.

Materials and methods
Patient population

The dataset used in this study consisted of a total of 102 NSCLC patients (Table 1). These patients underwent treatment at Dana Farber/Brigham Cancer Center between the years 2002 and 2014 with conventionally fractionated radiation therapy with curative intent. One pre-treatment CT scan and one post-treatment CT scan from each patient was included for analysis, totaling 204 scans. The patients were split using an 85%/15% split stratified on time-to-event to ensure an equal distribution of event times between the train and test sets resulting in 86 patients in the training dataset and 16 patients in the test dataset (Table 1). This study was approved by the Dana-Farber/Harvard Cancer Center Institutional Review Board (DF/HCC protocols 11–286 and 13–055) and conducted under a waiver of consent due to the retrospective nature of the study. All IRB guidelines were followed and this study was conducted in accordance with the principles of the Declaration of Helsinki.

Foundation model for cancer imaging biomarkers

The Foundation Model for Cancer Imaging Biomarkers¹⁰ (FMCIB) was trained using a modified version of the SimCLR framework¹¹ with a 3D ResNet50 encoder¹². For contrastive learning, positive examples were 50 × 50 × 50 mm³ 3D patches surrounding the lesion’s seed point and negative examples were 50 × 50 × 50 mm³ 3D patches randomly chosen from other areas of the scan. This model was trained on CT scans containing 11,467 annotated lesions from 2,312 unique patients. Additionally, this model demonstrated promising results for tasks including 2-year survival classification, diagnostics, anatomical awareness, prognosis, and test-retest which made it a good candidate for high-dimensional feature extraction. Pre-treatment and post-treatment 50 × 50 × 50 mm³ patches were input separately into this model. The outputs were feature vectors of size 4096 which were used in further analysis (Fig. 1).

Image preprocessing

CTs were acquired according to standardized scanning protocols at our institution using a GE “Lightspeed” CT scanner (GE Medical System, Milwaukee, WI, USA) for pre-treatment scans. Tumor segmentation was performed on all CTs using Eclipse software (Varian Medical System, Palo Alto, CA, USA) as previously described¹³. Localization of tumor regions was performed using seed-points defined as the center of mass of the tumor segmentations. To ensure compatibility with the FMCIB¹⁰, images were preprocessed to match the

Characteristic	All patients (n = 102)	Training dataset (n = 86)	Test dataset (n = 16)	p-value
Age (yrs)				
Median (IQR)	60.0 (9.3)	60.0 (9.4)	60.5 (8.7)	0.100
Sex				
Female	60 (58.8%)	49 (57%)	11 (68.8%)	0.380
Male	42 (41.2%)	37 (43%)	5 (31.2%)	0.380
Smoking status				
Median (IQR) pack years smoking	40.0 (27.9)	40.0 (30.8)	40.0 (36.5)	0.468
Tumor characteristics				
Clinical overall stage				
II (n = 9, 8.8%)	9 (8.8%)	9 (10.5%)	0 (0%)	0.175
A	2 (2.0%)	2 (2.3%)	0 (0%)	
B	7 (6.9%)	7 (8.1%)	0 (0%)	
III (n = 93, 91.2%)	93 (91.2%)	77 (89.5%)	16 (100%)	0.175
A	77 (75.5%)	65 (75.6%)	12 (75%)	
B	13 (12.7%)	9 (10.5%)	4 (25%)	
IV (n = 3, 2.9%)	3 (2.9%)	3 (3.5%)	0 (0%)	0.448
Median (IQR) clinical tumor size at diagnosis (cm)	4.2 (2.4)	4.4 (3.4)	3.3 (1.5)	0.054
Survival				
2-year survival	73 (72%)	62 (72%)	11 (68.8%)	0.786
Median (IQR) length of follow-up (yr)	3.64 (4.38)	4.07 (4.4)	2.54 (2.4)	0.26

Table 1. Characteristics of dataset stratified by training and test datasets. Variable distributions reported as n (%) unless otherwise specified. p-value calculated by two-sided t-test for continuous variables and chi-squared test for binary/categorical variables.

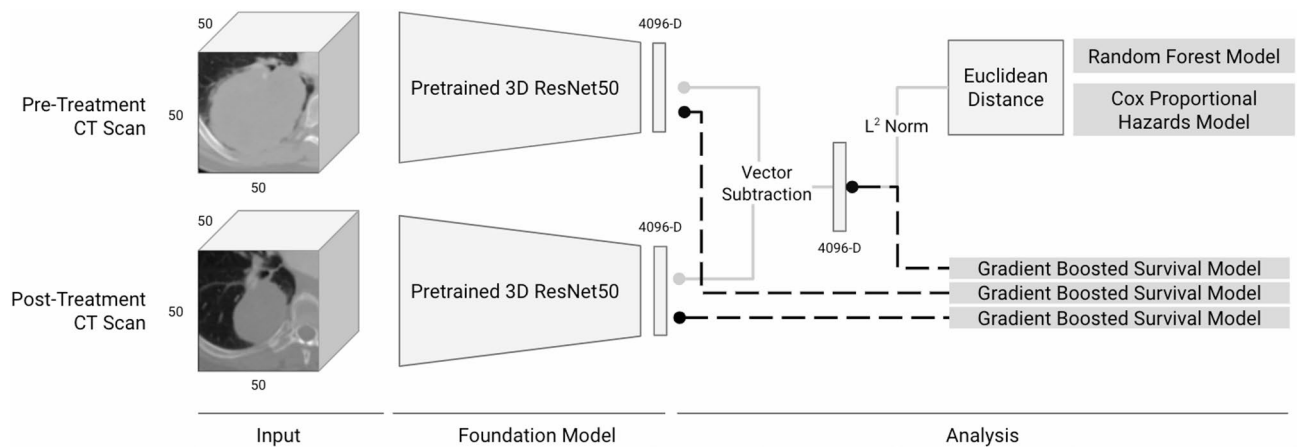


Fig. 1. Experimental design. FMCIB, with a 3D ResNet50 backbone and pre-trained on more than 11,000 lesions, is used in this study¹⁰. The foundation model is used on both pre-treatment and post-treatment CT scans to extract 4096-D feature vectors. The post-treatment feature vector is element-wise subtracted from the pre-treatment feature vector resulting in a single 4096-D vector. The element-wise subtraction vector is used to calculate a Euclidean distance that is passed along as a feature to a random forest clinical model and a Cox proportional hazards model. The element-wise subtracted vector alongside the feature vectors of the pre-treatment and post-treatment scans are also passed to a gradient boosting survival model.

model's training inputs. Specifically, $50 \times 50 \times 50$ mm³ volumes centered around the seed-points were then extracted and used as input to the model. CT voxels were resampled to $1 \times 1 \times 1$ mm³ using bilinear interpolation to address variability in slice thickness and in-plane resolutions. Voxel values were also normalized to a range of 0 and 1 using min-max normalization with the minimum being the lower-bound Hounsfield unit – 1024 and the maximum being the upper-bound Hounsfield unit 3072.

Feature-based analysis

The resultant 4096-D feature vectors output from FMCIB are a representation of the original localized tumor image data. Unlike hand-picked radiomic features, which tend to focus on specific aspects of the tumor, these foundation model features can often better capture additional information about the image as well as global spatial information. Moreover, these features avoid human subjectivity. As a result, and as shown by other studies, these features can be used out-of-the-box, without any finetuning, in a variety of tasks such as classification, regression, and survival tasks¹⁴. Furthermore, in addition to providing information about each individual image, because the input is a localized tumor that has been standardized, normalized, and always has the same orientation, it is expected that the features are directly comparable to one another. Thus, it is hypothesized that every entry between feature vectors, specifically between pre-treatment and post-treatment feature vectors, can be compared and that this comparison can capture the change between the original images. In order to test this property an element-wise vector subtraction was taken between the pre-treatment and post-treatment vectors. Before assessing the power of this new vector, exploratory analysis was done. The features were reduced to a dimensionality of 2 using principal component analysis (PCA) and visualized.

From here there were two main approaches taken to assessing the predictive power of the subtracted feature vector: (1) summarizing the vector using an L² norm thereby giving the Euclidean distance of the two original feature vectors and (2) using the subtracted feature vector in whole. These two forms of quantifying the difference between the CT scans were then used in 2-year survival classification models and time-to-event models. All analysis was performed in Python version 3.10 and all results were evaluated on the independent test dataset.

2-year survival classification

The Euclidean distance was taken between each pair of pre-treatment and post-treatment feature vectors. This distance was used as an additional feature in a 2-year survival random forest classification model also trained on baseline clinical features with known prognostic significance: age at time of radiation, gender, pack-years smoking, clinical tumor size at diagnosis (primary maximum axial diameter), and stage. The resultant model was compared to a clinical model trained without the Euclidean distance as an additional feature. 10-fold cross-validation was used during model training and the average area under the curve (AUC) for the hold-out set is reported.

Survival analysis

The Euclidean distance was also used as an additional feature alongside baseline clinical features in a Cox proportional hazards model and compared to a Cox proportional hazards model trained only on baseline clinical features. The models were compared by concordance index (C-index) and Akaike information criterion (AIC).

The feature vectors between pre-treatment and post-treatment scans were then subtracted element-wise and used as input to a gradient boosted survival model from scikit-survival (version 0.22.2)¹⁵. For comparison, 107 features were also extracted using the *RadiomicsFeatureExtractor* module from PyRadiomics (version 3.0.1) for both pre-treatment and post-treatment scans¹⁶. Similar to the FMCIB feature vectors, the PyRadiomics feature vectors were subtracted element-wise also for comparison. This analysis was compared to a gradient boosted survival model trained separately on: baseline clinical data, baseline clinical data with Euclidean distance, pre-treatment FMCIB feature vector only, and post-treatment FMCIB feature vector only. The gradient boosted survival models were evaluated using 1-year, 2-year, and 3-year cumulative AUC as well as the C-index. To further assess the robustness of the observed performance, we performed a permutation test in which the survival outcomes were randomly shuffled across patients in the test set, and the C-index and cumulative AUC were recomputed for each shuffle to generate an empirical null distribution. This distribution was then used to compute p-values for the observed metrics. Additionally, to compare between models, pairwise model comparisons were performed using a nonparametric bootstrap procedure (1,000 iterations). For each resample, cumulative AUCs were computed and two-sided empirical p-values were derived from the bootstrap distribution. The best-performing model at each time point (1, 2, and 3 years) was compared against the second- and third-best models as well as against the clinical baseline model.

Results

Clinical characteristics

The patient population consisted of 102 NSCLC patients (204 scans) treated with conventionally fractionated radiation therapy. Within the patient population 58.8% were female with a median age of 60 and predominantly diagnosed with stage III disease (91.2%) (Table 1). Overall, there was no statistically significant difference across the training and the test datasets for all patient parameters (Table 1). The median time from the pre-treatment CT scan to the start date of radiation therapy was 9 days with an IQR of 6 days. The median time between the pre-treatment CT scan and the post-treatment CT scan was 70 days with an IQR of 19 days. The median time from the end date of radiation therapy to the post-treatment CT scan was 18 days with an IQR of 15 days. The median follow-up time was 3.64 years with an IQR of 4.38 years (Table 1). Using the Kaplan Meier estimator, the median survival time was 4.91 years and survival estimates at 1, 2, and 3 years were 97.91%, 96.66% and 74.15% respectively.

Initial investigation of features

Taking the Euclidean distance between pre-treatment and post-treatment feature vectors across all patients in the study and analyzing these through a heatmap, we observe that matching pre-treatment and post-treatment vectors for the same patient tend to have a smaller Euclidean distance as expected (Fig. 2a). Visualizing small and large Euclidean distances, small Euclidean distances correspond to having little change in tumor appearance between pre-treatment and post-treatment CT scans, while large Euclidean distances represent a greater shift in appearance between scans (Fig. 2c). Further investigating the features output from the foundational model, we computed the element-wise difference between pre- and post-treatment feature vectors and applied PCA for dimensionality reduction. The first principal component was significantly correlated with clinical tumor size at diagnosis ($r=0.57$, $p < 1 \times 10^{-9}$) (Fig. 2b).

Gradient boosting survival analysis with foundation model derived features

Gradient boosted survival analysis was used to leverage the information coming from the entire feature vector in combination with time-to-event information. The baseline clinical model resulted in a C-index of 0.46 on the test dataset and improved to 0.53 when the Euclidean distance was added as a clinical feature (Supplementary Table 2). This same improvement is seen across the 1-year, 2-year, and 3-year cumulative AUCs as well as for clinical data without Euclidean distance compared to with (Table 2). Using just the pre-treatment and post-treatment FMCIB feature vectors independent of any clinical data resulted in C-indexes of 0.47 and 0.62 respectively (Supplementary Table 2), with the post-treatment feature vector also outperforming the clinical baseline model for 2-year and 3-year cumulative AUC (Table 2). Features extracted using PyRadiomics on both pre-treatment and post-treatment scans outperformed FMCIB pre-treatment and post-treatment features for both 1-year and 2-year cumulative AUC (Table 2). However, these results were not statistically significant. Training the gradient boosted survival model on the element-wise subtracted FMCIB feature vector resulted in the highest C-index on the test dataset of 0.74 and 1-year and 2-year cumulative AUCs of 0.93 and 0.91 respectively with the 2-year cumulative AUC achieving statistical significance ($p=0.006$, Table 2). Performing a dimensionality reduction using PCA demonstrated a similar result (Supplementary Table 3).

The best performing model at each year was then compared pairwise to the second- and third-best performing models and also pairwise compared to the model trained on clinical features. At 1-year, the element-wise model outperformed the clinical model ($p=0.004$) but showed no significant difference compared to the model trained on PyRadiomics post-treatment features ($p=0.25$) or PyRadiomics pre-treatment features ($p=0.078$). At 2 years, the element-wise model again outperformed the clinical model ($p=0.0040$) and the model trained on FMCIB post-treatment features ($p=0.010$), while no significant difference was observed when compared to the model trained on PyRadiomic post-treatment features ($p=0.16$). By 3 years, the best performing model was the model trained on FMCIB post-treatment features which had no significant differences found between any comparator (element-wise $p=0.39$; FMCIB pre-treatment $p=0.40$; clinical $p=0.40$).

Euclidean distance as a clinical feature

A random forest clinical model incorporating baseline clinical data (age at time of radiation, gender, pack-years of smoking, clinical tumor size at diagnosis, and stage) yielded an AUC of 0.52. In comparison, when

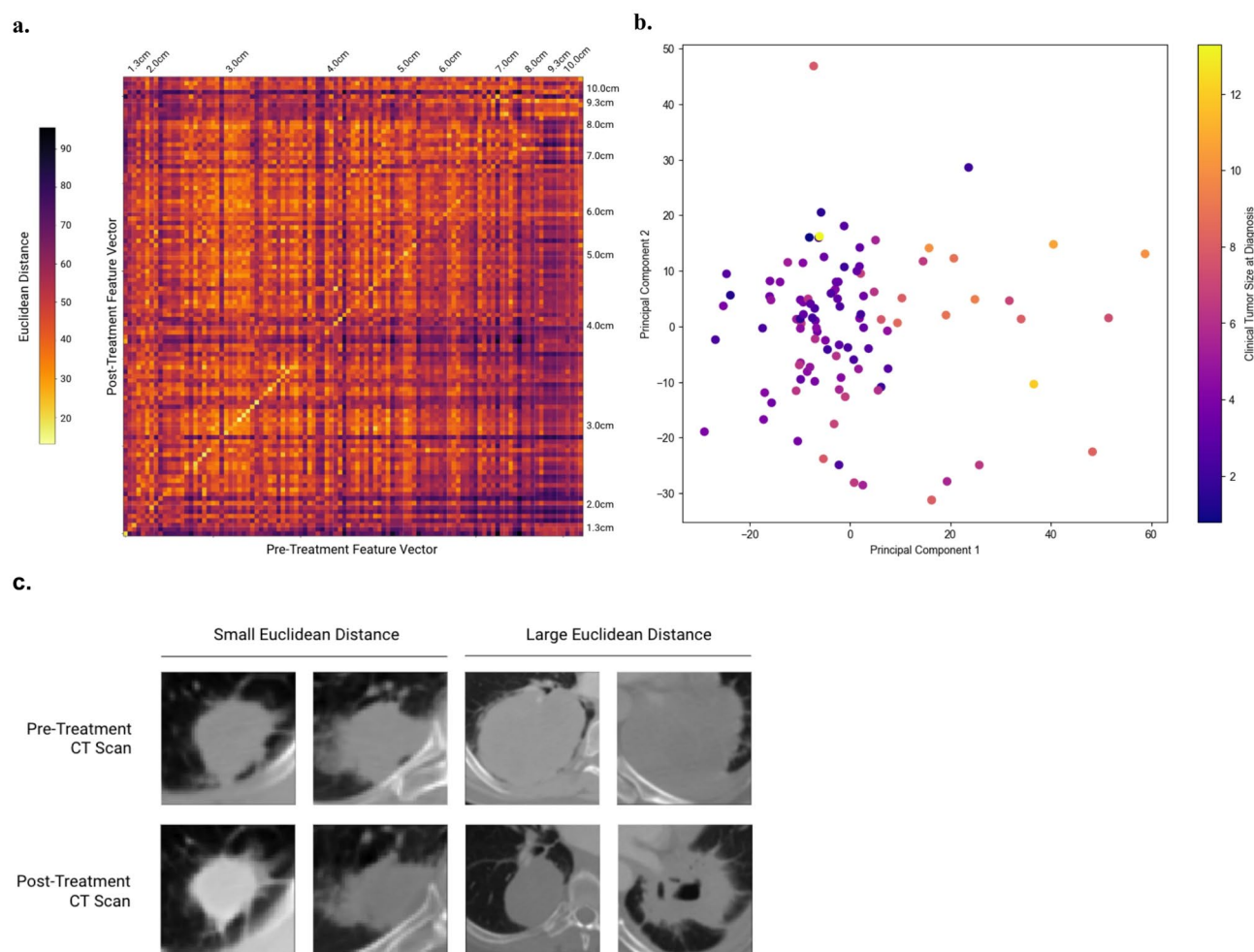


Fig. 2. Investigation of features. **(a)** Heatmap of Euclidean distance between pre-treatment feature vectors and post-treatment vectors where the patients are organized the same on both axes by increasing tumor size. The smaller Euclidean distances (light yellow) along the diagonal of the heatmap demonstrates that for most patients the Euclidean distance is the smallest when comparing their pre-treatment scan to their respective post-treatment scan. **(b)** PCA dimensionality reduction of the element-wise subtracted feature vector for all patients, colored by clinical tumor size at diagnosis demonstrating a correlation between clinical tumor size at diagnosis and the element-wise subtracted features. **(c)** Pre-treatment and post-treatment CT scans for the two patients with the smallest Euclidean distance between respective feature vectors and the two patients with the largest Euclidean distance between respective feature vectors demonstrating that a smaller Euclidean distance correlates to tumors that visually changed less between scans.

the Euclidean distance between the 4096-D feature vectors was added to the random forest clinical model as a feature, the model demonstrated a general improvement in predictive power with the AUC increasing to 0.59. The Euclidean distance showed no significant correlation with any of the other baseline clinical features in the data (Supplementary Fig. 1).

In addition to a random forest clinical model, survival analysis was conducted. A Cox proportional hazards model fit to the same baseline clinical data resulted in a C-index of 0.59 on the training dataset, a C-index of 0.59 on the test dataset, and an overall AIC of 364.24. In this model none of the coefficients were statistically significant. The Cox proportional hazards model slightly decreased in its predictive power when the Euclidean distance feature was added resulting in a C-index of 0.61 on the training dataset, a C-index of 0.51 on the test dataset, and an overall AIC of 365.66.

Discussion

We demonstrate that adapting a previously developed CT-based foundation model (FMCIB) for feature extraction combined with serial CT scans to quantify feature change in response to treatment can result in the identification of novel prognostic biomarkers despite a small clinical dataset. Moreover, this work identified the element-wise subtracted feature vector as being the most predictive of survival.

Other studies similarly have shown the power of extracting meaningful features from medical imaging data for predicting clinical outcomes^{17–19}. Further research has demonstrated the ability of deep-learning derived

Training features	Cumulative AUC (p-value)		
	1-year	2-year	3-year
Clinical data without euclidean distance	0.33 (0.68)	0.47 (0.53)	0.53 (0.53)
Clinical data with euclidean distance	0.4 (0.61)	0.55 (0.36)	0.56 (0.35)
PyRadiomics features for pre-treatment scans	0.73 (0.39)	0.39 (0.74)	0.33 (0.82)
PyRadiomics features for post-treatment scans	0.8 (0.25)	0.73 (0.11)	0.35 (0.80)
Element-wise subtracted PyRadiomics features	0.66 (0.68)	0.5 (0.39)	0.33 (0.80)
FMCIB features for pre-treatment scans	0.06 (0.94)	0.39 (0.58)	0.65 (0.49)
FMCIB features for post-treatment scans	0.33 (0.93)	0.58 (0.71)	0.75 (0.17)
Element-wise subtracted FMCIB features	0.93 (0.13)	0.91 (0.0060)	0.68 (0.14)

Table 2. 1-year, 2-year, and 3-year cumulative AUCs and p-values for the gradient boosted survival model trained on different feature sets. FMCIB features are compared to clinical data and features extracted from PyRadiomics. Bolded are the best performing feature sets across each cumulative AUC.

radiomic features to be even more powerful than more traditional, manually extracted features. But behind all of this work there remain two major caveats (1) most deep-learning models are analyzed at a single time-point and (2) in order to be effective, these models require large datasets which, in the field of medical imaging, is often not possible. This study aimed to understand the ability of medical imaging foundation models to produce clinically relevant features without the need for any additional training as well as to understand if these features are robust enough to extract information across time.

Using the element-wise subtracted feature vector in a gradient boosted survival analysis addressed both of these limitations by demonstrating the capability for the foundation model to produce clinically-relevant, predictive features across time points. Specifically, across 1-year and 2-year survival prediction, this element-wise subtracted feature vector showed the strongest predictive power, including outperforming the model trained strictly on clinical features. Additionally, at 2-years, the element-wise model significantly outperformed the model trained only on post-treatment FMCIB features. By 3 years, however, the predictive differences between models were no longer significant, suggesting that imaging-derived features may be most informative for early outcome prediction, whereas longer-term survival may be driven by non-imaging factors such as heterogeneity in treatment response or tumor biology.

While adding Euclidean distance to the baseline random forest clinical model predicting 2-year survival did not produce a statistically significant result, the improvement in the model's AUC as well as the decrease in the overall p-value points towards the potential predictive power of the Euclidean distance as a clinical feature. Although this result did not hold for the survival analysis using the Cox proportional hazards model, it did hold for the gradient boosted survival analysis where, again, there was an increase in all four metrics being tested (1-year, 2-year, and 3-year cumulative AUCs and C-index) when the Euclidean distance was added as a feature to baseline clinical data.

Beyond survival prediction, our exploratory PCA analysis further underscored the clinical relevance of the extracted features. Specifically, the first principal component of the element-wise subtracted feature vectors was significantly correlated with tumor size at diagnosis ($r=0.57$, $p<1\times10^{-9}$). This association demonstrates that the foundation model-derived features capture information related to well-established clinical measurements such as tumor size, while the imperfect correlation indicates that they also encode additional, non-size-based characteristics of tumor biology. Importantly, this suggests that the proposed approach can reveal prognostically meaningful information that is not captured by traditional dimensional assessments such as RECIST, highlighting the added value of foundation model features in advancing beyond conventional imaging biomarkers.

In addition to requiring no additional deep-learning training or fine tuning, the FMCIB features were extracted using single seed point tumor localization and did not require full tumor segmentations. CT scans are already a part of routine clinical workflow and passing them through a model such as this one requires very little additional manual work other than identifying the seed point of the tumor.

Previous works have also demonstrated the importance of deep learning within the field of medical imaging for predicting clinical outcomes. More specifically, several studies have focused on single time point feature extraction to predict clinical outcomes^{20–24}. While these works show promising results, they are limited by only having access to information about a patient's state at one point in time. Fewer studies have focused on leveraging medical imaging from multiple time points. The dataset used in this work overlaps with the dataset introduced in a study by Xu et al., which similarly used a pretrained model and seed point tumor localization to extract features from multiple time points to predict survival²⁵. While the work from Xu et al. also shows the importance of having follow-up scans, it required fine tuning the ResNet backbone for feature extraction and training an RNN for survival prediction²⁵. Moreover, many of these studies use a backbone pretrained on ImageNet which is limiting because medical imaging data has little resemblance to the images in ImageNet. This study thus further differs by leveraging a model that has been pretrained on medical imaging data.

Despite the strengths of this study, there remain several limitations, the largest of which was sample size. The foundation model was trained on a larger dataset, thereby making the features more robust and reliable, but further analysis using the extracted features has decreased power due to small sample size. The analysis was also completed only on the primary tumor and not on the nodal disease which was present in many of the patients in the dataset. Moreover, no external validation was done for this work. Another major limitation is our

constrained ability to understand and interpret the high dimensional feature vectors representing the images. In comparison to a more concrete value such as the RECIST diameter, each entry in a 4096-D vector, on its own, has little meaning. While previous studies have used activation maps to better understand the outputs of deep-learning models, these remain difficult to extract and not always meaningful²¹.

To address these limitations as well as expand beyond this work, there are several possibilities for future directions. This includes training these same models on a larger dataset and testing the work on an external dataset. It would also be interesting to incorporate more time points beyond just pre-treatment and post-treatment, test the predictive power of these models on other outcomes, and expand to other treatment modalities such as immunotherapy.

In conclusion, this work demonstrates two major findings: (1) foundation models can be effective in extracting clinically relevant features without the need for a large dataset and (2) changes in features across multiple time points can be predictive of survival. Moreover, these changes in features tend to outperform models based solely on clinical data and can outperform models based on single time point imaging features. Overall, using foundation models and serial data point analysis over-time can improve the predictive ability of deep learning-based models for clinical applications.

Data availability

The training dataset is internal to Mass General Brigham institutions and contains sensitive protected health information. Due to privacy concerns and legal restrictions associated with patient data, the complete dataset cannot be made publicly available. Researchers interested in accessing the dataset can submit a formal request detailing the intended use of the data to R.H.M. (RMAK@partners.org). Each request will be evaluated on a case-by-case basis in compliance with the ethical guidelines and agreements under which the data were collected.

Received: 4 July 2025; Accepted: 28 October 2025

Published online: 03 December 2025

References

1. Siegel RL, Giaquinto AN & Jemal, A. Cancer Statistics, 2024. *CA Cancer J. Clin.* **74** (1). <https://doi.org/10.3322/caac.21820> (2024).
2. Ganti, A. K., Klein, A. B., Cotarla, I., Seal, B. & Chou, E. Update of Incidence, Prevalence, Survival, and initial treatment in patients with Non-Small cell lung cancer in the US. *JAMA Oncol.* **7** (12), 1824–1832. <https://doi.org/10.1001/jamaoncol.2021.4932> (2021).
3. Eisenhauer, E. A. et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur. J. Cancer.* **45** (2), 228–247. <https://doi.org/10.1016/j.ejca.2008.10.026> (2009).
4. Zhou, T. et al. The effectiveness of RECIST on survival in patients with NSCLC receiving chemotherapy with or without target agents as First-Line treatment. *Sci. Rep.* **5** (1), 7683. <https://doi.org/10.1038/srep07683> (2015).
5. Litière, S. et al. The components of progression as explanatory variables for overall survival in the response evaluation criteria in solid tumours 1.1 database. *Eur. J. Cancer.* **50** (10), 1847–1853. <https://doi.org/10.1016/j.ejca.2014.03.014> (2014).
6. William, W. N. et al. Computed tomography RECIST assessment of histopathologic response and prediction of survival in patients with resectable non-small-cell lung cancer after neoadjuvant chemotherapy. *8* (2), 222–228. <https://doi.org/10.1097/jto.0b013e3182774108> (2013).
7. Gulstene, S. et al. What is the predictive value of RECIST criteria following stereotactic lung radiation? *Radiation Oncol.* **190**, 109976. <https://doi.org/10.1016/j.radonc.2023.109976> (2024).
8. Ma, J. et al. Segment anything in medical images. *Nat. Commun.* **15** (1). <https://doi.org/10.1038/s41467-024-44824-z> (2024).
9. Alzubaidi, L. et al. *MedNet: Pre-Trained Convolutional Neural Network Model for the Medical Imaging Tasks.* <https://arxiv.org/pdf/2110.06512.pdf>.
10. Pai, S. et al. Foundation model for cancer imaging biomarkers. *Nat. Mach. Intell.* **6** (3), 354–367. <https://doi.org/10.1038/s42256-024-00807-9> (2024).
11. Chen, T., Kornblith, S., Norouzi, M. & Hinton, G. A simple framework for contrastive learning of visual representations. **30**. <https://arxiv.org/abs/2002.05709> (2020).
12. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. **10**, <https://arxiv.org/abs/1512.03385> (2015).
13. Coroller, T. P. et al. Radiomic-Based pathological response prediction from primary tumors and lymph nodes in NSCLC. *J. Thorac. Oncol.* **12** (3), 467–476. <https://doi.org/10.1016/j.jtho.2016.11.2226> (2017).
14. Notley, S. & Magdon-Ismael, M. Examining the use of neural networks for feature extraction: A comparative analysis using deep learning, support vector machines, and K-nearest neighbor classifiers. <https://doi.org/10.48550/arXiv.1805.02294>
15. Polsterl, S. scikit-survival: A library for Time-to-Event analysis built on top of scikit-learn. *J. Mach. Learn. Res.* **21**, 1–6 (2020). <https://www.jmlr.org/papers/volume21/20-729/20-729.pdf>
16. van Griethuysen, J. J. M. et al. Computational radiomics system to Decode the radiographic phenotype. *Cancer Res.* **77** (21), e104–e107. <https://doi.org/10.1158/0008-5472.CAN-17-0339> (2017).
17. Lao, J. et al. A deep Learning-Based radiomics model for prediction of survival in glioblastoma multiforme. *Sci. Rep.* **7**, 10353. <https://doi.org/10.1038/s41598-017-10649-8> (2017).
18. Sun, R. et al. Preoperative CT-based deep learning radiomics model to predict lymph node metastasis and patient prognosis in bladder cancer: a two-center study. *Insights Imaging.* **15**, 21. <https://doi.org/10.1186/s13244-023-01569-5> (2024).
19. Lian, J. et al. Predicting progression-free survival in patients with epithelial ovarian cancer using an interpretable random forest model. *Heliyon* **10**, e35344. <https://doi.org/10.1016/j.heliyon.2024.e35344> (2024).
20. Huang, B. et al. Prediction of lung malignancy progression and survival with machine learning based on pre-treatment FDG-PET/CT. *eBioMedicine* **82** <https://doi.org/10.1016/j.ebiom.2022.104127> (2022).
21. Chaunzwa, T. L. et al. Deep learning classification of lung cancer histology using CT images. *Sci. Rep.* **11** (1). <https://doi.org/10.1038/s41598-021-84630-x> (2021).
22. Kumar, D., Wong, A. & Clausi, D. A. Lung nodule classification using deep features in CT images. *IEEE Xplore* <https://doi.org/10.1109/CRV.2015.25>.
23. Lao, J. et al. A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. *Sci. Rep.* **7** (1). <https://doi.org/10.1038/s41598-017-10649-8> (2017).
24. Haarburger, C., Weitz, P., Rippel, O. & Merhof, D. Image-based survival prediction for lung cancer patients using CNNs. *IEEE Xplore* <https://doi.org/10.1109/ISBL.2019.8759499>.
25. Xu, Y. et al. Deep learning predicts lung cancer treatment response from serial medical imaging. *Clin. Cancer Res.* **25** (11), 3266–3275. <https://doi.org/10.1158/1078-0432.CCR-18-2495> (2019).

Acknowledgements

The authors acknowledge a philanthropic gift from George Denny supporting FH and JP.

Author contributions

The concept for this study was developed and designed by J.P., S.P., Y.X., and R.M. Acquisition and interpretation of data were done by J.P., J.H., F.H., Y.X., D.C., H.A., and R.M. Code and reproducibility were the responsibility of J.P. The paper was written by J.P. Critical revision of the paper for important intellectual content was carried out by J.P., S.P., J.H., F.H., Y.X., D.C., R.M., H.A. All authors provided final approval of the manuscript and agree to be accountable for all aspects of the work. The study was supervised by R.M.

Funding

We acknowledge financial support from the National Institute of Health (NIH) (Grant nos. NIH (NCI) 5U01CA209414, NIH-USA U24CA194354, NIH-USA U01CA190234, NIH-USA U01CA209414, NIH-USA R35CA22052).

Declarations

Competing interests

RHM: Advisory Board (ViewRay, AstraZeneca), Consulting (AstraZeneca, Varian Medical Systems, Sio Capital Management), Honorarium (Novartis, Springer Nature), Research Funding (National Institute of Health, ViewRay, AstraZeneca, Siemens Medical Solutions USA, Inc, Varian Medical Systems). HJWLA reports consulting fees from Onc.AI, Love Health, and Sphera, and stock from Onc.AI, Sphera, Love Health, Health-AI, Ambient, and AstraZeneca, outside the submitted work. All other authors do not have any competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-26365-7>.

Correspondence and requests for materials should be addressed to R.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025