

Novel Designs of Radiomics-Integrated Deep Learning Models

by

Zongsheng Hu

Graduate Program in Medical Physics  
Duke Kunshan University and  
Duke University

Date: \_\_\_\_\_

Approved:

\_\_\_\_\_  
Chunhao Wang, Supervisor

\_\_\_\_\_  
Fang-Fang Yin

\_\_\_\_\_  
David Huang

Thesis submitted in partial fulfillment of  
the requirements for the degree of  
Master of Science in the  
Graduate Program in Medical Physics of  
Duke Kunshan University and  
Duke University

2022

ABSTRACT

Novel Designs of Radiomics-Integrated Deep Learning Models

by

Zongsheng Hu

Graduate Program in Medical Physics  
Duke Kunshan University and  
Duke University

Date: \_\_\_\_\_

Approved:

\_\_\_\_\_  
Chunhao Wang, Supervisor

\_\_\_\_\_  
Fang-Fang Yin

\_\_\_\_\_  
David Huang

An abstract of a thesis submitted in partial  
fulfillment of the requirements for the degree  
of Master of Science in the  
Graduate Program in Medical Physics of  
Duke Kunshan University and  
Duke University

2022

Copyright by  
Zongsheng Hu  
2022

## **Abstract**

Purpose:

To investigate the feasibility of integrate radiomics and deep learning in computer-aided medical imaging analysis

Methods:

Two different approaches were investigated to integrate radiomics and deep learning on two independent tasks respectively. In the first approach, a 2D sliding kernel was implemented to map the impulse response of radiomic features throughout the entire chest X-ray image; thus, each feature is rendered as a 2D map in the same dimension as the X-ray image. Based on each of the three investigated deep neural network architectures, including VGG-16, VGG-19, and DenseNet-121, a pilot model was trained using X-ray images only. Subsequently, 2 radiomic feature maps (RFMs) were selected based on cross-correlation analysis in reference to the pilot model saliency map results. The radiomics-boosted model was then trained based on the same deep neural network architecture using X-ray images plus the selected RFMs as input.

The proposed radiomics-boosted design was developed using 812 chest X-ray images with 262/288/262 COVID-19/Non-COVID-19 pneumonia/healthy cases, and 649/163 cases were assigned as training-validation/independent test sets. For each model, 50 runs were trained with random assignments of training/validation cases following

the 7:1 ratio in the training-validation set. Sensitivity, specificity, accuracy, and ROC curves together with Area-Under-the-Curve (AUC) from all three deep neural network architectures were evaluated.

In the second approach, a cohort of 235 GBM patients with complete surgical resection was divided into short-term/long-term survival groups with 1-yr survival time threshold. Each patient received a pre-surgery multi-parametric MRI exam with 4 scans: T1, contrast-enhanced T1 (T1ce), T2, and FLAIR. Three tumor subregions were segmented by neuroradiologists, and the whole dataset was divided into training, validation, and test groups following a 7:1:2 ratio.

The developed model comprises three data source branches: in the 1<sup>st</sup> radiomics branch, 456 radiomics features (RF) were calculated from the three tumor subregions of each patient's MR images; in the 2<sup>nd</sup> deep learning branch, an encoding neural network architecture was trained for survival group prediction using each single MR modality, and high-dimensional parameters from the last two network layers were extracted as deep features (DF). The extracted radiomics features and deep features were processed by a feature selection procedure to reduce the dimension size of each feature space. In the 3<sup>rd</sup> branch, patient-specific clinical features (PSCF), including patient age and three tumor subregions volumes, were collected from the dataset. Finally, data sources from all three branches were fused as an integrated input for a supporting vector machine (SVM) execution for survival group prediction. Different strategies of model design

were investigated in comparison studies, including 1) 2D/3D-based image analysis, 2) different radiomics feature space dimension reduction methods, and 3) different data source combinations in SVM input design.

Results:

In the first approach, all three investigated deep neural network architectures demonstrated improved sensitivity, specificity, accuracy, and ROC AUC results in COVID-19 and healthy individual classifications. VGG-16 showed the largest improvement in COVID-19 classification ROC (AUC from 0.963 to 0.993), and DenseNet-121 showed the largest improvement in healthy individual classification ROC (AUC from 0.962 to 0.989). The reduced variations suggested improved robustness of the model to data partition. For the challenging Non-COVID-19 pneumonia classification task, radiomics-boosted implementation of VGG-16 (AUC from 0.918 to 0.969) and VGG-19 (AUC from 0.964 to 0.970) improved ROC results, while DenseNet-121 showed a slight yet insignificant ROC performance reduction (AUC from 0.963 to 0.949). The achieved highest accuracy of COVID-19/Non-COVID-19 pneumonia/healthy individual classifications were 0.973 (VGG-19)/0.936 (VGG-19)/ 0.933 (VGG-16), respectively.

In the second approach, the model achieved 0.638 prediction accuracy in the test set when using patient-specific clinical features only, which was higher than the results using radiomics features/deep features as sole input of SVM in both 2D and 3D based

analysis. The inclusion of radiomics features or deep features with patient-specific clinical features improved accuracy results in 3D analysis. The most accurate models in 2D/3D analysis reached the highest accuracy of 0.745 with different combinations of dissimilarity-selected radiomics features, deep features, and patient-specific clinical features, and the corresponding ROC area-under-curve (AUC) results were 0.69 (2D) and 0.71 (3D), respectively.

#### Conclusions:

The integration of radiomic analysis in deep learning model design improved the performance and robustness computer-aided diagnosis and outcome predication, which holds great potential for clinical applications and provides a radiomics perspective for deep learning interpretation.

# Contents

Abstract .....	iv
List of Tables .....	x
List of Figures .....	xi
1. Introduction .....	1
1.1 Imaging based prognosis and diagnosis .....	1
1.2 Radiomics based medical imaging analysis .....	2
1.3 Deep learning based medical imaging analysis .....	3
1.4 The integration of radiomics and deep learning .....	5
1.5 Literature review .....	7
1.5.1 COVID-19 diagnosis .....	7
1.5.2 Survival prediction of GBM patient .....	9
2 Materials and Methods .....	12
2.1 COVID-19 and non-COVID-19 pneumonia classification .....	12
2.1.1 Image Dataset .....	12
2.1.2 Neural Network Architecture .....	12
2.1.3 Radiomic Feature Map Extraction .....	15
2.1.4 Model Training and Evaluation .....	16
2.2 Glioblastoma post-resection survival prediction .....	17
2.2.1 Image Dataset .....	17
2.2.2 Model Design .....	18

2.2.2.1 Radiomics Analysis.....	19
2.2.2.2 Deep Feature Extraction.....	20
2.2.2.3 SVM Design .....	22
2.2.3 Comparison Studies.....	23
3. Result .....	24
3.1 COVID-19 and non-COVID-19 pneumonia classification.....	24
3.2 Glioblastoma Post-Resection Survival Prediction .....	31
4. Conclusion .....	36
4.1 COVID-19 and non-COVID-19 pneumonia classification.....	36
4.2 Glioblastoma post-resection survival prediction.....	41
References .....	45

## List of Tables

Table 1. Sensitivity, specificity, accuracy, and ROC AUC results summaries of top: VGG-16 architecture, middle: VGG-19 architecture, bottom: DenseNet-121 architecture. .... 27

Table 2. Survival group prediction results of deep neural network output as in Figure 4. .... 32

Table 3. Survival group prediction results of SVM implementation with single data source branch in Figure 1. DF: deep feature; RF: radiomics feature; TS: tumor size ..... 33

Table 4. Survival group prediction results of SVM implementation with multiple data source branches in Figure 1. DF: deep feature; RF: radiomics feature; TS: tumor size..... 34

## List of Figures

Figure 1. Diagrams of the three studied deep neural network. (A) VGG-16, (B) VGG-19 and (C) DenseNet-121 .....	14
Figure 2. A workflow summary of radiomic feature map (RFM) calculation in this work. ....	15
Figure 3. A flowchart illustration of the radiomics-integrated deep learning model. ....	19
Figure 4. Diagrams of the studied deep neural network: a. VGG-16 architecture for 2D based analysis. b. The encoding part of U-net for 3D based analysis. ....	22
Figure 5. Image comparisons from 3 example cases for the VGG-16 pilot model. The GLRLM SRE RFMs and saliency map (overlaid with X-ray image) are illustrated in 0.3 power scale. ....	25
Figure 6. The ROC results of pilot model vs radiomics-boosted model using (a) VGG-16; (b). VGG-19; and (c). DenseNet-121 deep neural network architecture.....	29
Figure 7. The SM cross-correlation matrix of radiomics-boosted model on test set for left: VGG-16; middle: VGG-19; right: DenseNet121 architectures.....	31
Figure 8. The ROC curves from the identified multi-dimensional deep learning model with the best accuracy results in 2D (blue) and 3D (red) scenarios. ....	34
Figure 9. SVM coefficients of features in 2D (left) and 3D (right) models with the best prediction accuracy results. ....	35

# 1. Introduction

## 1.1 *Imaging based prognosis and diagnosis*

Accurate and efficient prognosis and diagnosis are crucial to the treatment and management of patients. The prognosis and diagnosis enable personalized medicine and treatment strategy, which spares the limited medical resources while enhances the life quality of patients. Besides the symptoms and screening test, medical imaging is a promising non-invasive approach for prognosis and diagnosis<sup>1-3</sup>. Depends on different imaging modalities, the radiographic phenotype can show both anatomical (such as CT, MRI) and metabolic (such as PET, SPECT, fMRI) information of patients, which provides a reliable basis for the diagnosis and prognosis of the disease. However, currently diagnosis and prognosis are done manually by radiologists, including the examination of medical images to identify the possible disease locations, which is highly dependent on the experiences of the radiologist. In addition, when dealing with high patient volume with large medical dataset, this process is laborious and hardly consistent<sup>4,5</sup>

Therefore, computer-aided medical imaging analysis (CAMIA) system is highly demanded to solve the aforementioned problem with high throughput quantitative analysis. Previous studies revealed that CAMIA systems outperformed radiologists in many challenging tasks<sup>6,7</sup>. Furthermore, with CAMIA information as reference information, radiologist reading results could be significantly improved<sup>8-10</sup>. Therefore,

the development of accurate and efficient CAMIA system has been being an important research topic.

## **1.2 Radiomics based medical imaging analysis**

One approach for such CAMIA system is radiomics-based image analysis, which extracts radiomics features as computation image biomarkers. These features are calculated using hand-craft mathematical algorithms to reveal important information that hard to be appreciated by radiologist<sup>11-13</sup>. Numerous radiomics features has been developed to describe different characteristic of medical imaging. For instance, some describe the size and shape information<sup>14</sup>, some are based on the image intensity analysis<sup>15</sup>, the widely used gray-level co-occurrence matrix (GLCOM), gray-level run length matrix (GLRLM) describe the relationship between image voxels<sup>16,17</sup>, some uses mathematic transformation to process the image<sup>18</sup>. These radiomics features were widely used in a wide range of application.

One important application is treatment evaluation and prognosis. King et al<sup>19</sup> found that in human patients with head-and-neck cancer, the average apparent diffusion coefficient histogram kurtosis, and skewness on diffusion weighted MRI images were found to be higher in poorly responded tumors during chemoradiotherapy. Another study<sup>20</sup> constructed multi input models to predict pathologic response to chemoradiotherapy using a PET images dataset from 20 esophageal cancer patients. They found the combination of different types of radiomics features significantly

improved the model performance than using single types of radiomics feature. Coroller et al<sup>21</sup> constructed a model to predict distant metastasis and survival in 182 lung cancer patients using 35 and 6 CT radiomics features and achieve a promising performance.

Radiomics analysis is also widely used in tissue identification task. As early as 1990s, GLCOM textures on T1 and T2-weighted MR images were found to have potential to differentiate brain tumor, edema, cerebrospinal fluid (CSF), white matter, and gray matter, in patients with brain cancer<sup>22</sup>. Petkovska et al<sup>23</sup> constructed a radiomics based classification model to identify malignant and benign nodules from contrast-enhanced CT images. GLCOM textures were calculated, and the model achieve a higher performance than radiologists in the classification task. A more recent study<sup>24</sup> combines CT and PET-based textures to classify malignant and benign lesion in various tumor sites and achieved an accuracy over 0.7.

### ***1.3 Deep learning based medical imaging analysis***

Deep learning models composed of many layers that transform input data to outputs while learning increasingly higher-level features, which enable computers to automatically extract the features that optimally represent the image information for the problem at hand. In imaging analysis task, the mostly used model are convolutional neural networks (CNNs) which is firstly proposed in 1980<sup>25</sup> and applied to medical imaging analysis in 1995<sup>26</sup>. Driven by recent theoretical developments and accesses to

massive computation power, deep learning has demonstrated its great potential in CAMIA.

In recent studies, deep learning-based algorithms were widely investigated in computer aided diagnosis. Hosseini-Asl et al<sup>27</sup> constructed 3D CNN based on MRI scan of the brain to classify patients into cognitively normal older individuals, cognitive impairment, and Alzheimer's disease, achieving an accuracy of 89.1%. Liu et al<sup>28</sup> proposed a multi-view multi-scale CNN to classify the lung nodule into 6 types (i.e., well-circumscribed, vascularized, juxta-pleural, pleural-tail, ground glass optical, and non-nodules) in CT scans. The proposed method achieves an overall accuracy of 92.1% with a dataset comprising 1012 sample. In order to deal with the limited size of dataset, Khan et al<sup>29</sup> using the concept of transfer learning in breast cancer detection and classification based on breast cytology images. Pre-trained GoogleNet, VGGNet, and ResNet were used to extract features from images. These features were then fed into fully connected layer for classification of malignant and benign cells, achieving an overall classification accuracy of 97.5%.

Another important application of deep learning-based CAMIA is segmentation. The purpose of segmentation is to identify voxels in medical images which can shape the contour of the region of-interest (ROI). Accurate segmentation is crucial in treatment planning for improving the treatment outcome of patients. Li et al<sup>30</sup> proposed a novel hybrid densely connected U-Net consisting of a 2D Dense-U-Net and a 3D counterpart

for liver and tumor segmentation, achieving DICE of 0.96 and 0.82 for liver and lesion respectively. Wang et al<sup>31</sup> developed a patient-specific adaptive CNN that was trained on past weekly MRI to segment lung tumors on current weekly MRI. Data augmentation was used to address the limited size of dataset. The proposed framework achieves a precision and DICE of 0.81 and 0.82 respectively.

#### ***1.4 The integration of radiomics and deep learning***

Although radiomics and deep learning achieve promising performance in CAMIA, both methods have their own limitations. The radiomics analysis usually use hand-crafted features that are carefully designed and manually extracted based on previous studies or prior knowledge. Although such feature extraction is straightforward, these feature descriptors are unable to automatically adapt to the problem at hand. In addition, while radiomics features can render important information from medical images, many radiomics features show poor reproducibility and robustness. The value of radiomics features is heavily dependent on acquisition modes, matrix size, reconstruction algorithms, iteration numbers, image noise, and artifacts<sup>32-34</sup>. Unlike classic radiomics analysis that requires handcrafted feature extraction with experimental knowledge, deep learning models calculate a very large number of features simultaneously and automatically extract important features for the problem at hand. Nevertheless, like all other deep learning applications in medical image analysis, the hyperparameters in the neural network are generated without

explicit human knowledge intervention. Thus, the 'black box' nature of deep learning-based CAMIA inhibits their interpretability, and potential clinical applications of these CAMIAs could be impaired by limited interpretability by clinicians. Furthermore, the overfitting problem of deep learning is tricky when dealing with medical imaging dataset with limited samples.

Therefore, the integration of radiomics analysis and deep learning implementation for quantitative image analysis is conceptually appealing: on one hand, quantitative information extracted by deep neural network can potentially work jointly with radiomics features in a complementary way to improve risk modelling performance; on the other hand, potential interactions between radiomics features as handcrafted descriptors with known expressions and complex deep features may provide a radiomics perspective for deep learning model explanation to address well-known 'black-box' issue. In this work, two different approaches were investigated to integrate radiomics and deep learning on two independent tasks respectively. In the first approach, we included the radiomic features in the input of deep learning model. In this approach, we aim to develop a radiomics-boosted deep learning CAD design for chest X-ray based COVID-19 diagnosis. An innovative implementation of radiomics analysis was included to analyze deep features from 3 custom-trained neural networks for COVID-19, non-COVID-19 pneumonia, and healthy individuals' classification, and such radiomics analysis results were then incorporated as additional image-based input

source for 3 improved neural network designs. The proposed methodology may enhance deep learning interpretability for COVID-19 diagnosis from a current radiomics knowledge perspective. In the second approach, we establish a novel multi-dimensional deep learning design that enables radiomics analysis integration. It features two innovative solutions: 1) classic radiomics features and deep features, i.e., latent variables of feature extraction layers in a deep neural network, are extracted simultaneously from a multi-parametric MRI protocol; and 2) radiomics features and relevant non-image clinical features in lower dimensions are connected to the deep features in an abstract feature space, and such connected feature is utilized by Support Vector Machine (SVM)<sup>35</sup> as a supervised machine learning model for GBM survival risk group prediction. We demonstrate our development using a publicly available GBM image dataset, and we conducted multiple comparison studies to study its potential merits.

## ***1.5 Literature review***

### **1.5.1 COVID-19 diagnosis**

Since its first discovery in 2019, the coronavirus disease (COVID-19) has affected more than 100 million people globally, and more than 6 million deaths related to COVID-19 were reported by 15<sup>th</sup> March 2022<sup>36</sup>. Accurate and efficient diagnosis of COVID-19 is crucial to interrupt disease transmission and to start treatments of affected individuals. Currently, reverse transcription-polymerase chain reaction (RT-PCR) has been recognized as the gold standard for COVID-19 diagnosis for its high specificity<sup>37</sup>.

While RT-PCR test may have limited sensitivity and longtime of processing (a few hours to 2 days)<sup>38</sup>, radiographic procedures, including chest X-ray and CT exams, have been adopted clinically as alternative diagnosis tools<sup>39</sup>. While COVID-19 related abnormalities could be more easily found in volumetric CT images<sup>40</sup>, planar chest X-ray has its unique advantages in COVID-19 diagnosis. Specifically, the short imaging time on a more accessible X-ray unit enables rapid COVID-19 exams, which can be critical in areas with high-volume patients and/or limited-resource medical facilities. To date, pilot studies have revealed that certain X-ray image features, including peripheral consolidations and ground-glass opacities, have been widely observed in COVID-19 infected patients<sup>39-44</sup>. However, the prevalent application of chest X-ray imaging in COVID-19 diagnosis is challenged by relatively limited sensitivity and specificity<sup>40</sup>. Additionally, radiographic exams including chest X-ray may not be optimal for radiologists' reading in the differentiation of non-COVID-19 pneumonia from COVID-19, which is important for early patient stratification that can lower COVID-19 mortality rate with more targeted treatments<sup>5,6</sup>.

Computed aided diagnosis systems (CAD) may have the potential to solve the problem with high throughput quantitative analysis. In the last several months under the COVID-19 pandemic, studies revealed that CAD systems outperformed radiologists in radiographic-based COVID-19 diagnosis; with CAD information as reference information, radiologist reading results could be significantly improved<sup>5,9,45</sup>. One

approach for such CAD system is radiomics-based image analysis, which first extracts radiomics features as computation image biomarkers and then uses the extracted features in hand-made or machine learning classifier tasks. Although handcrafted radiomics features are commonly used in medical image analysis with possible qualitative image interpretability, the reported accuracy (75-80%) of COVID-19 diagnosis is still limited in the representative radiomics-based CAD works<sup>46-48</sup>. Driven by recent theoretical developments and accesses to massive computation power, deep learning has demonstrated its great potential in CAD developments. It has been reported that deep learning solution based on artificial neural network deployment could achieve high (>90%) specificities in COVID-19 diagnosis against healthy individuals<sup>49,50</sup>; moreover, decent specificities (>85%) of differentiating COVID-19 from non-COVID-19 pneumonia have been achieved<sup>5,51-53</sup>.

### **1.5.2 Survival prediction of GBM patient**

Arising from glial cells, glioblastoma (GBM) is the most common primary malignant brain tumor, accounting for 16% of all central nervous system neoplasms<sup>54</sup>. The average age-adjusted incidence rate of GBM is 4.43 per 100,000 population<sup>55</sup>. In the meantime, GBM is one of the most fatal cancers with short overall survival (OS): over 70% of GBM patients will suffer disease progression within one year after diagnosis<sup>56</sup>, with a five-year survival rate of less than 5%<sup>55</sup>.

For individualized patient management in clinical practice, accurate and reliable predictions of GBM patient survival are highly demanded to enable early treatment interventions and alternative strategy exploration; however, such survival prediction remains challenging. Studies showed that GBM patients with the same tumor histopathology may have significantly different survivals, which renders the histopathologic approach questionable<sup>57</sup>. Radiography analysis, particularly MRI-based image analysis, aims at decoding radiographic phenotype of tumor and normal brain from high soft tissue contrast at millimeter level, and thus becomes an active field for GBM OS prediction<sup>58</sup>. In particular, quantitative MR image analysis via computational approach, represented by radiomics techniques, extracts high-level descriptive statistics beyond typical image rendering dimensions. These statistics were hypothesized to benefit GBM survival prediction in combination with advanced statistical modelling<sup>16,17</sup>. For instance, a representative study by Kickingreder et al extracts radiomics features from multi-parametric anatomical MRI exams for supervised principal analysis, and this approach outperformed classic risk model in low risk/high risk patient group stratification<sup>59</sup>. Other radiomics studies have studied quantitative MRI protocol use, genetic profile integration<sup>60</sup>, and novel statistical approaches<sup>61</sup> for enhanced GBM survival prediction. Recently, deep learning (DL) approaches have been investigated for MR-based GBM OS prediction. Unlike classic radiomics analysis that requires handcrafted feature extraction with experimental knowledge, deep learning models

calculate a very large number of features simultaneously, while the risk modelling via deep neural network is supported by massive computations without explicit expressions<sup>62</sup>. To date, pilot works have reported GBM survival group stratification and regression based on deep learning designs<sup>63-65</sup>.

## **2 Materials and Methods**

### **2.1 COVID-19 and non-COVID-19 pneumonia classification**

#### **2.1.1 Image Dataset**

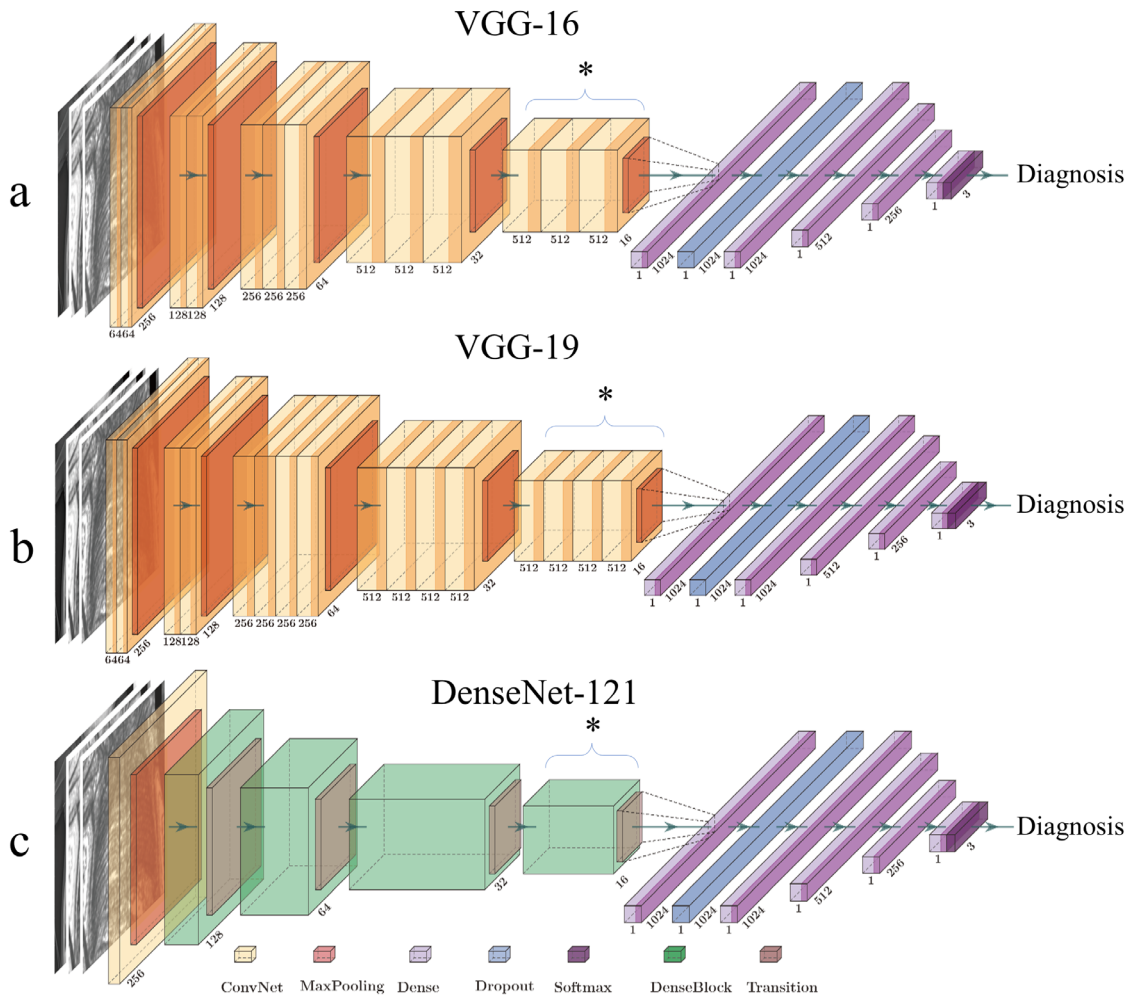
In this IRB-waived retrospective study, a total of 812 chest X-ray images were collected from three public databases<sup>66-68</sup>, including 262/288/262 images of COVID-19/Non-COVID-19 pneumonia/healthy individuals, respectively. The image numbers from the three categories were approximately the same for eliminating categorical bias during deep learning training. All collected images were verified by experienced medical physicists with proper lung X-ray display settings and with absences of overlaid image reading annotations. To unify image data size, all images were resized to a 256x256 matrix grid size using b-spline interpolation and were normalized to 256 gray levels. 649 and 163 images (8:2) were assigned for the model training set and the independent test set, respectively.

#### **2.1.2 Neural Network Architecture**

To investigate radiomics-boosted deep learning as a general methodology, we studied 3 deep neural networks, VGG-16 (Figure 1 (A)), VGG-19 (Figure 1(B)), and DenseNet-121 (Figure 1(C)), for COVID-19/Non-COVID-19 pneumonia/healthy individual classification. Based on pre-trained deep learning schemes<sup>69,70</sup>, the investigated 3 deep neural network share a same two-part design: the 1<sup>st</sup> part is convolutional base: In VGG-16 and VGG-19, the convolutional bases consist of 5

convolutional blocks. Each convolutional block is stacked by 2 to 4 convolutional layers and a max pooling layer. In each convolutional layer, the filter size is 3×3 with padding and stride of 1. Max-pooling is performed over a 2×2-pixel window with a stride of 2. In DenseNet-121, the convolutional base consists of one convolutional block and 4 dense block. The convolutional block consists of 7×7 convolutional layers with a stride of 2 and a 3×3 Max-pooling with a stride of 2. The Dense Blocks consists of 6, 12, 24, and 16 convolutional unites, respectively. Each unite are stack by an 1×1 and a 3×3 convolutional layers. The first 3 dense blocks are followed by 3 transition layers, which consist of a 1×1 convolutional layers with stride 1 and a 2×2 average pool with stride 2; the 2<sup>nd</sup> part is the Dense part, which is the stack of Dense layers. Depending on specific classification tasks, the number and size of Dense layers can be customized. In all three deep neural network architectures, the self-defined Dense classifier connects convolutional base and consists of five Dense layers with the size of 1024, 1024, 512, 256, and 3, respectively. The input of neural network is a three-channel image with a 256×256×3 shape size, while the output is one of the three categorical binary label vectors, i.e., [1,0,0], [0,1,0], and [0,0,1], which correspond to COVID-19, non-COVID-19 pneumonia, and healthy results, respectively. To deal with relatively small data size in this work, the convolutional bases loaded the weights that were pre-trained on ImageNet as a transfer learning scheme<sup>71</sup>. In addition, in order to make the models more relevant for the problem at hand, fine-tune technique was used with the last blocks

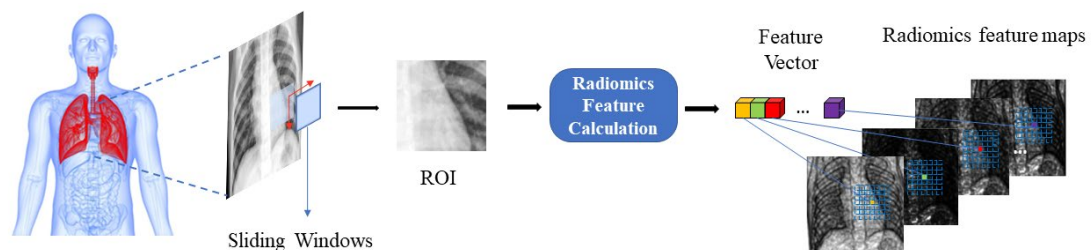
(marked \* in Figure 1: the last convolutional blocks in VGG-16 and VGG-19, as well as the last dense block in DenseNet121) being set as free parameters for task-specific training. To avoid the occurrence of overfitting, a dropout layer was added between the first two Dense layers with a dropout possibility of 0.5, and soft-max activation was used in the output layer.



**Figure 1. Diagrams of the three studied deep neural network. (A) VGG-16, (B) VGG-19 and (C) DenseNet-121**

## 2.1.3 Radiomic Feature Map Extraction

Classic radiomics analysis calculates radiomic features as scalar values from a pre-defined region-of-interest (ROI) in image space. While this approach has been widely adopted to capture the overall textures in ROI, it cannot capture the anatomy-driven subtle texture variations within the ROI. As such, we implement a radiomic feature map calculation workflow<sup>72</sup>, which is summarized in Figure 2. For RFM generation, a 2D kernel (13x13 matrix size) was adopted to form a region-of-interest (ROI), and 37 radiomic features were extracted as a 1x37 vector within this ROI following classic GLCOM<sup>16</sup> (21 features) and GLRLM<sup>17</sup> (16 features) feature extraction methods using 32 grey levels. For each feature, the calculated feature value was assigned as the pixel value centered at the ROI. By moving this 2D kernel across the X-ray image as a sliding window operation, 37 feature maps were formed in the same dimension as original X-ray images. All radiomic analysis was done using custom code that was benchmarked with digital phantoms<sup>73</sup> and complies with the imaging biomarker standardization initiative<sup>74</sup>.



**Figure 2. A workflow summary of radiomic feature map (RFM) calculation in this work.**

## 2.1.4 Model Training and Evaluation

For each of the three investigated deep neural network architecture, 2 model versions were trained: in the 1<sup>st</sup> pilot model, the X-ray image is the sole model input. To accommodate the input shape of pre-trained neural networks, the grayscale X-ray images were broadcast to three channels as a network input variable. This model serves as the benchmarking deep learning model in this work; in the 2<sup>nd</sup> version as a radiomics-boosted model, the grayscale X-ray image and two derived RFMs were stacked as the 3-channel neural network input variable. These two RFMs were selected based on the analysis of the 1<sup>st</sup> model's saliency map (SM), which indicates how important each pixel is with respect to the final classification results of the neural network in the benchmarking model. It is calculated as the absolute gradient of class activation which is defined as the dot product of prediction output and target divided by the input image<sup>75</sup>: a pixel with a higher intensity value in SM indicates higher importance of that pixel in neural network's attention for diagnosis. The two RFMs with the highest average cross-correlation (CC) values against the SM results in training data were selected. This action amplifies certain pixels (and regions) with potentially high importance of disease diagnosis in the image space, which could improve the overall diagnostic accuracy of the proposed radiomics-boosted model in comparison with the 1<sup>st</sup> pilot model.

To investigate model robustness, 50 runs of each model version were trained using the training data set (649 images). In each run, the training and validation samples were randomly selected following a 7:1 ratio. During deep learning training within the TensorFlow environment using a Nvidia™ (Santa Clara, CA) Tesla V100 graphic card, the loss function was categorical cross-entropy, and Adam optimizer was selected. For model evaluation, sensitivity, specificity, accuracy, and ROC area under the curve (AUC) results from both model versions were analyzed. Statistical significance of comparison was determined by Wilcoxon signed rank tests at level 0.05.

## ***2.2 Glioblastoma post-resection survival prediction***

### **2.2.1 Image Dataset**

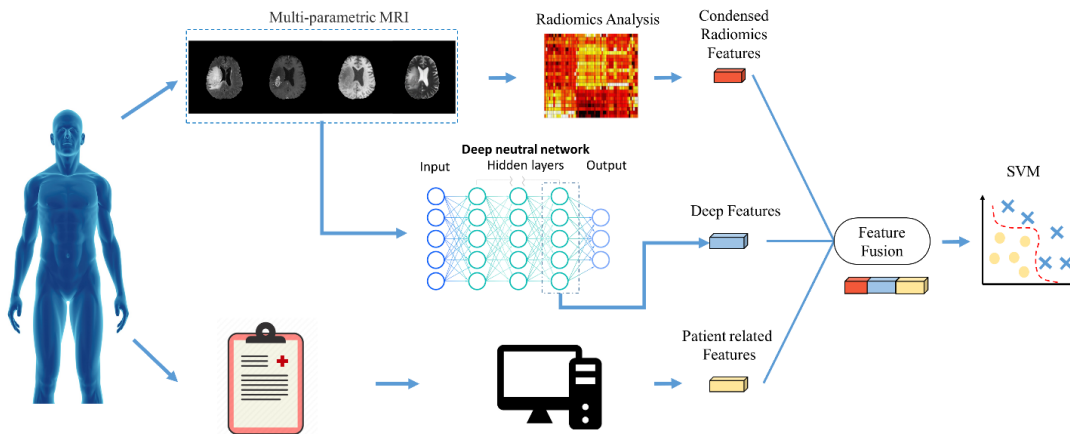
In this study, we studied a total of 235 GBM patients from the BraTS 2020 database<sup>76,77</sup>. Each patient has a pathologically confirmed diagnosis and a pre-operative multi-parametric MRI exam that included four series: T1 weighted, T2 weighted, contrast-enhanced T1 weighted (T1ce), and water-suppressed FLAIR (Fluid Attenuation Inversion Recovery). All four MR series have been co-registered with skull-skipping processing, and image resolutions have been unified to isotropic 1 mm<sup>3</sup>. In addition to survival in days following complete resection surgery, each patient has ground-truth tumor segmentation results from experienced neuro-radiologist contouring. These segmentation results include three key regions: contrast-enhanced tumor region,

peritumoral edema, and the necrotic/non-enhancing tumor core. In this work, these patients were divided into two groups: 116 patients in the short-term survival group (<1yr survival) and 119 patients in the long-term survival group (>1yr survival) after complete resection surgery. Such group assignment follows most clinical relevance for a balanced data group assignment.

### **2.2.2 Model Design**

Figure 3 summarizes the overall design of our radiomics-integrated deep learning workflow for GBM survival group prediction. As illustrated, multi-dimensional data processing comprises three branches. In the 1<sup>st</sup> branch, radiomics features are calculated using each MR modality from selected regions-of-interest (ROIs), and dimension reduction methods are adopted to select condensed radiomics features from the whole feature space. In the 2<sup>nd</sup> branch, an encoding neural network architecture is trained using each single MR modality for short-term/long-term survival prediction. Higher level features from last two dense layers are extracted as deep features, followed by dimension reduction processing. In the 3<sup>rd</sup> branch, the patient-specific clinical features (PSCF) were collected from the dataset. Due to the public dataset content, we collected patient age at the time of surgery and calculated 3 tumor region volumes (i.e., contrast-enhanced region, non-enhancing region, and peritumoral edema) based on ground-truth segmentation results. Finally, data from all three branches are fused as an integrated input for SVM modelling to predict the survival group. In this work, we

investigated different strategies of the designed model implementation, including 1) 2D/3D approaches of image analysis; 2) different dimension reduction strategies; and 3) different combinations of SVM input design from the three branches. The whole patient dataset was divided into training, validation, and test set following a ratio of 7:1:2.



**Figure 3. A flowchart illustration of the radiomics-integrated deep learning model. A total of 3 data branches are included, top: radiomics features; middle: deep features; bottom: patient-specific clinical features were extracted and compressed as the fused input of SVM**

### 2.2.2.1 Radiomics Analysis

In this work, we extracted 22 textural features from Gray Level Co-Occurrence Matrix (GLCOM)<sup>16</sup> and 16 textural features from Gray Level Run Length Matrix (GLRLM)<sup>17</sup> using our in-house radiomics software package<sup>73,78</sup>, and features from three segmentation regions, i.e., contrast-enhanced tumor region, peritumoral edema, and the necrotic tumor core, were collected. In 2D based analysis, the axial slice with the biggest cross-sectional area of the contrast-enhanced tumor region was selected, and radiomics

features were extracted from the three segmented regions in the selected slice. In 3D based analysis, radiomics features were extracted from the 3D segmentation of the regions with isotropic resolution. Therefore, 456 radiomics features ((22+16) features x 4 MR modalities x 3 regions) were derived from each patient in 2D/3D analysis. It is a known issue that radiomics features may demonstrate high correlations with each other<sup>79</sup>. As such, proper dimension reduction methods are in demand to reduce the influence of repetitive information and overfitting possibilities. In this work, we studied two approaches in feature selection: principal component analysis (PCA) and dissimilarity analysis. In the PCA approach, 14 main components were obtained for maintaining >90% explain ratio of original full radiomics feature space. In dissimilarity analysis, radiomics features that maximized differences between long-term and short-term survival groups were selected based on the defined dissimilarity below:

$$Dissimilarity = \frac{|mean(L) - mean(S)|}{|mean(L)| + |mean(S)|}$$

where mean(S) and mean(L) refer to the mean values of the feature over short-term and long-term survival patient groups, respectively. 7 radiomics features were selected with dissimilarity over the threshold of 0.15.

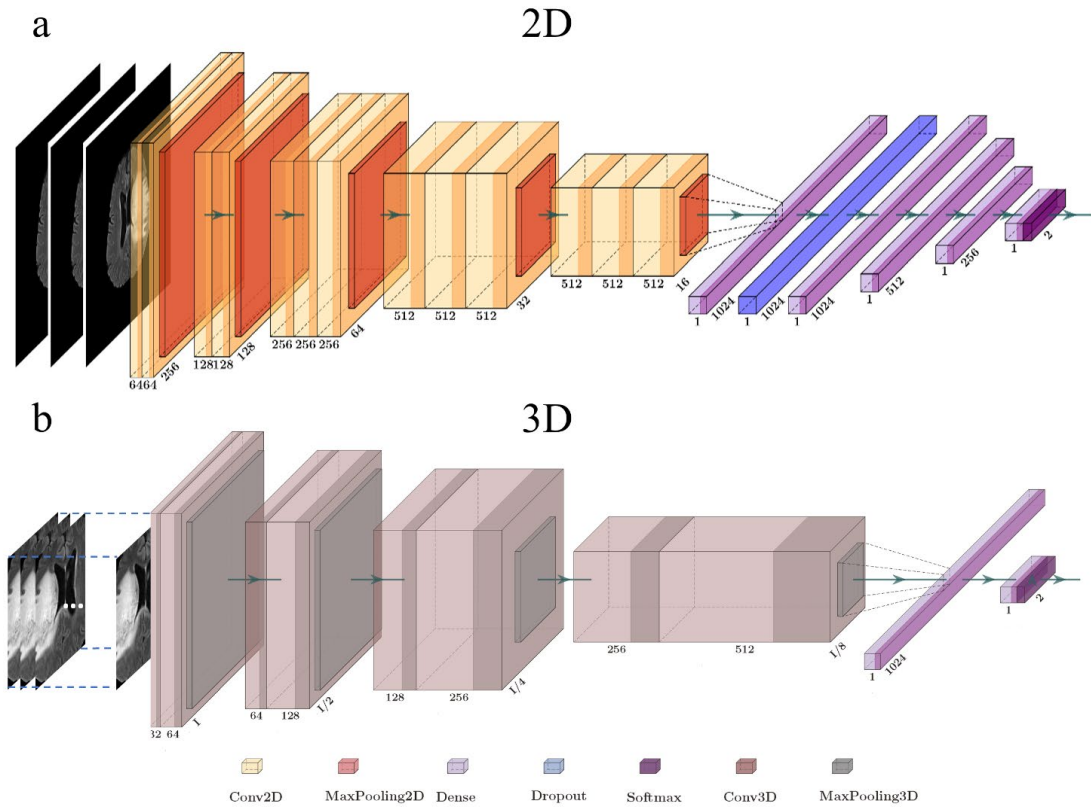
#### **2.2.2.2 Deep Feature Extraction**

In this work, we adopted different deep neural network designs in the 2<sup>nd</sup> branch of Figure 3 to accommodate 2D and 3D based image analysis. Figure 4A summarizes the

deep neural network for 2D image inputs. Based on a pre-trained VGG-16 architecture<sup>69</sup>, the network consists of two parts: the 1<sup>st</sup> part is a convolutional base with 5 convolutional blocks. Each convolutional block is stacked by 2 or 3 convolutional layers and a max-pooling layer. In each convolutional layer, the filter size is 3×3 with padding and stride of 1. Max-pooling is performed over a 2×2-pixel window with a stride of 2. The 2<sup>nd</sup> part is the dense part, which is the stack of 5 dense layers with the size of 1024, 1024, 512, 256, and 3, respectively. To avoid the occurrence of overfitting, a dropout layer was added between the first two Dense layers with a dropout possibility of 0.5, and soft-max activation was used in the output layer. The input of the neural network is a three-channel image with a 160×208×3 shape size following the BraTS dataset specification, while the output is categorical binary label vectors, i.e., [1,0] and [0,1], which correspond to short-term and long-term survival groups, respectively. To deal with relatively small data size in this work, the convolutional base loaded the weights that were pre-trained on ImageNet<sup>80</sup> as a transfer learning scheme<sup>81</sup>.

Figure 4B summarizes the deep neural network for 3D image inputs. Similarly, the proposed 3D CNN architecture consists of a convolutional base and dense part. The convolutional base is constructed based on the encoding part of U-Net architecture<sup>82</sup>. The convolutional base is loaded with the weights that are pre-trained with the segmentation task on a medical imaging dataset<sup>83,84</sup>. The convolutional base consists of 4 convolutional blocks. Each convolutional block is stacked by 2 or 3 3D convolutional

layers and a 3D max-pooling layer. In each convolutional layer, the filter size is  $3 \times 3 \times 3$  with padding and stride of 1. Max-pooling is performed over a  $2 \times 2 \times 2$ -pixel window with a stride of 2. The dense part is the stack of 2 dense layers with a size of 1024 and 2. Soft-max activation was used in the output layer.



**Figure 4. Diagrams of the studied deep neural network: a. VGG-16 architecture for 2D based analysis. b. The encoding part of U-net for 3D based analysis.**

### 2.2.2.3 SVM Design

The extracted features in Figure 3 were concatenated as a fused feature, an SVM model was trained to predict short-term/long-term survival group results. Prior to SVM training, all integrated inputs were normalized to  $[-1, 1]$  range. During SVM training, a

linear kernel was selected to enable studies of input components' contribution to final prediction results. The C value, which represents misclassifying tolerance during training, was set to  $40^{85}$ .

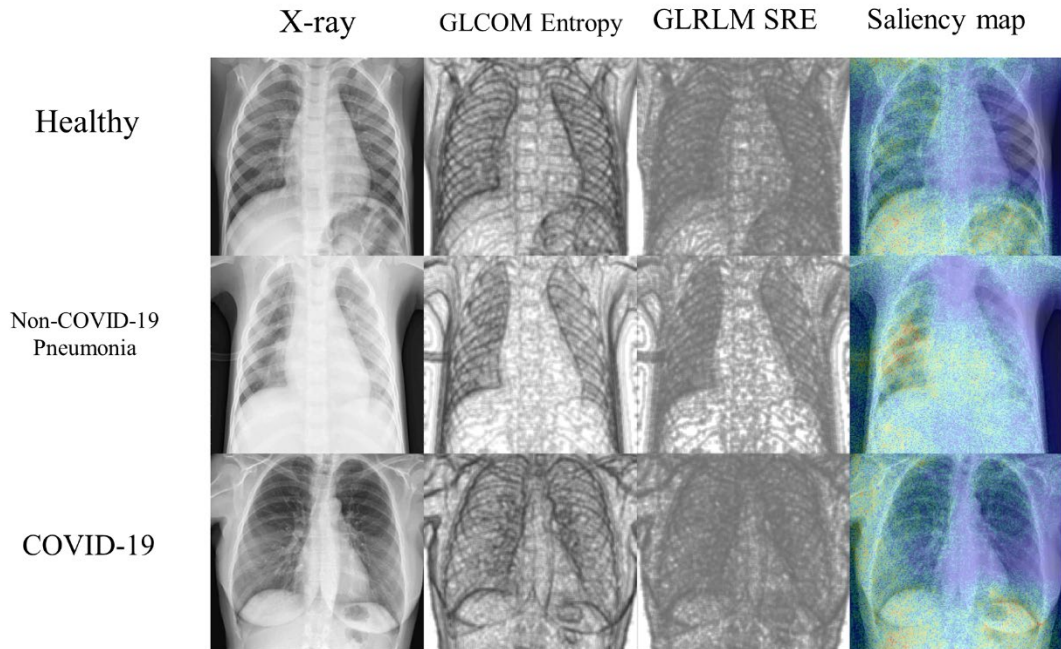
### **2.2.3 Comparison Studies**

We evaluated the performance of the proposed deep learning design in a two-fold comparison study design. First, we compare GBM survival group prediction accuracies using sole data source, i.e., deep learning network results, radiomics modelling results, and PSCF modelling results. A same SVM design was used for radiomics modelling and PSCF modelling. As the next step, we compared prediction accuracies of our proposed deep learning model with different radiomics dimension reduction strategies and different integrated feature designs of SVM input in Figure 3. Both 2D and 3D scenarios were studies.

### **3. Result**

#### ***3.1 COVID-19 and non-COVID-19 pneumonia classification***

Figure 5 shows an example of image comparison in the pilot model of VGG-16, i.e., the one using X-ray only as input. As illustrated, the identified RFMs, GLCOM Entropy (CC = 0.33) RFM, and GLRLM Short Run Emphasis (SRE) (CC = 0.31), render more tissue textural variations in both lung and other soft tissue regions than original X-ray images. Similarly, GLCOM Entropy (CC = 0.32) and GLCOM Sum Entropy (CC = 0.31) RFMs were selected from VGG-19 pilot model, while the GLCOM Sum Average (CC = 0.27) and GLRLM Short Run High Gray Level Emphasis (SRHGLE) (CC = 0.28) RFMs were selected from DenseNet-121 pilot model.



**Figure 5. Image comparisons from 3 example cases for the VGG-16 pilot model. The GLRLM SRE RFMs and saliency map (overlaid with X-ray image) are illustrated in 0.3 power scale.**

Table 1 summarizes the quantitative comparisons of sensitivity, specificity, accuracy, and ROC AUC between pilot models and radiomics-boosted models. For VGG-16 architecture, the radiomics-boosted deep learning model achieved statistical improvements in all parameters with  $p < 0.05$ . The largest improvements were observed in non-COVID-19 pneumonia diagnosis. Additionally, the reduced standard deviations of the reported statistics indicated the enhanced robustness of the radiomics-boosted deep learning design. These quantitative results highlight the superiority of the proposed radiomics-boosted deep learning model in the context of VGG-16 architecture. The standard deviation of 50 runs is less than 3%, reaching a high level of robustness.

For VGG-19 architecture, the mean values of all parameters were higher in the radiomics-boosted model than the pilot model; however, the observed numerical improvements were small, and only a few improvements were found with statistical significance in COVID-19 and healthy class results. The standard deviations were also reduced as indicators of improved model robustness. Results from DenseNet-121 architecture were similar to VGG-16 and VGG-19 results except for non-COVID-19 pneumonia classification, in which mixed impacts in the radiomics-boosted model were presented; nevertheless, the radiomics-boosted model did improve COVID-19 and healthy individual classification performance and increased model robustness in the healthy individual classification. As a summary, the radiomics-boosted design achieved best performance in COVID-19 diagnosis in both VGG-16 and VGG-19 architecture applications, while it achieved the best performance in healthy individual classification in DenseNet-121 architecture application.

**Table 1. Sensitivity, specificity, accuracy, and ROC AUC results summaries of top: VGG-16 architecture, middle: VGG-19 architecture, bottom: DenseNet-121 architecture. The mean values and standard deviation of 50 trained runs are reported.**

VGG-16	Healthy		Non-COVID-19 pneumonia		COVID-19	
	X-ray	X-ray+RFM	X-ray	X-ray+RFM	X-ray	X-ray+RFM
Sensitivity	0.854±0.065	0.922±0.059*	0.780±0.092	0.857±0.0361*	0.903±0.071	0.949±0.036*
Specificity	0.918±0.044	0.938±0.022*	0.941±0.041	0.963±0.023*	0.940±0.037	0.973±0.020*
Accuracy	0.895±0.029	0.933±0.023*	0.892±0.029	0.931±0.016*	0.927±0.028	0.965±0.016*
AUC	0.948±0.027	0.979±0.012*	0.918±0.043	0.969±0.017*	0.963±0.023	0.993±0.006*

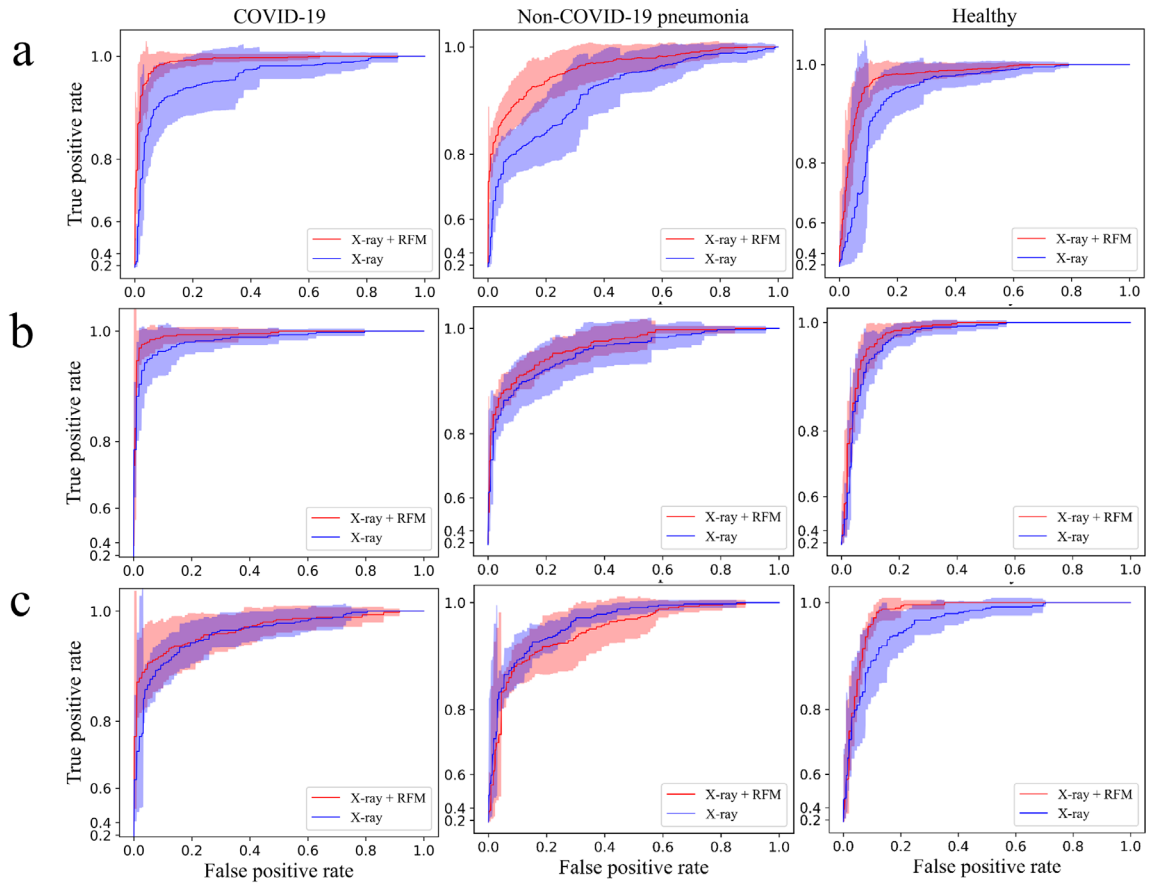
VGG-19	Healthy		Non-COVID-19 pneumonia		COVID-19	
	X-ray	X-ray+RFM	X-ray	X-ray+RFM	X-ray	X-ray+RFM
Sensitivity	0.902±0.097	0.921±0.053	0.824±0.059	0.854±0.041	0.933±0.047	0.969±0.039*
Specificity	0.920±0.030	0.940±0.013*	0.969±0.017	0.972±0.020	0.969±0.028	0.975±0.023
Accuracy	0.914±0.032	0.933±0.019	0.925±0.025	0.936±0.013	0.956±0.023	0.973±0.017*
AUC	0.970±0.015	0.978±0.007	0.964±0.028	0.970±0.013	0.987±0.013	0.994±0.006*

DenseNet-121	Healthy		Non-COVID-19 pneumonia		COVID-19	
	X-ray	X-ray+RFM	X-ray	X-ray+RFM	X-ray	X-ray+RFM
Sensitivity	0.810±0.210	0.948±0.048*	0.902±0.070	0.872±0.042	0.747±0.150	0.800±0.061
Specificity	0.920±0.049	0.924±0.045*	0.850±0.109	0.922±0.036*	0.971±0.037	0.994±0.014*
Accuracy	0.881±0.056	0.920±0.020*	0.866±0.060	0.907±0.017*	0.896±0.038	0.928±0.024*
AUC	0.962±0.022	0.989±0.008*	0.963±0.017	0.949±0.022	0.964±0.022	0.972±0.027

Figure 6 summarizes the ROC analysis results of the three studied architectures.

The blue and red solid lines represent the average ROC results of 50 runs of two deep model versions (X-ray only vs. X-ray + RFM), and the colored bands represent the model performance variation as  $\pm 1$  standard deviation. For VGG-16 architecture (Figure 6a), radiomics-boosted design improved ROC results of all three classification tasks, and the

largest performance improvement was observed in non-COVID-19 pneumonia diagnosis. Additionally, the radiomics-boosted deep learning model has narrower ROC bandwidth, which suggests the enhanced robustness of its design under different data sample uses. The same improvements were also observed in VGG-19 results (Figure 6b), but the improvements' magnitudes were smaller than ones in VGG-16 results, which was mainly contributed by the higher performance of VGG-19 prior to radiomics-boosted implementation. In Figure 6c of DenseNet-121 results, while the ROC improvement by radiomics-boosted design was prominent in healthy individual classification, the improvement in COVID-19 diagnosis was limited. As reported in Table 1, the ROC result in non-COVID-19-pneumonia showed a slightly decreased performance after radiomics-boosted design, though such decrease was found no statistical significance.



**Figure 6. The ROC results of pilot model vs radiomics-boosted model using (a) VGG-16; (b). VGG-19; and (c). DenseNet-121 deep neural network architecture. 0.3 power scale was used in the y axis to highlight the difference.**

As an example of saliency map (SM) visualization, the SM results of the VGG-16 architecture are illustrated in the last column of Figure 5. The pixel values in SM can be interpreted as the attention of the deep learning model. As seen, the attention patterns, i.e., colored hot regions distribution in SMs, demonstrated prominent spatial heterogeneity across the image field-of-view. In addition, the current SM illustration suggests potential class-specific spatial patterns of deep network attention: more

attention might be drawn to lateral lung regions for COVID-19 detection, while mediastinum regions might be attention focus for non-COVID-19 pneumonia detection. In order to quantitatively analyze the attention patterns across different patient cohorts, we calculated the CC matrix for SMs of all three radiomics-boosted architectures in the test set in Figure 7, which includes all the CCs between paired SM results from the test set. For VGG-16 and VGG-19 results, CCs within each cohort were relatively higher than those calculated across different cohorts. This result suggests that the developed deep learning model captured cohort-specific features for the classification task. Additionally, the mean CC result of COVID-19 *vs* non-COVID-19 pneumonia (VGG-16:0.12; VGG-19: 0.10) cohort was slightly higher than the result of COVID-19 *vs* Healthy cohorts (VGG-16:0.07; VGG-19: 0.08) and non-COVID-19 pneumonia *vs* Healthy cohorts (VGG-16:0.09; VGG-19: 0.09). This observed COVID-19/non-COVID-19 pneumonia similarity supports the clinical reports of challenges in COVID-19/non-COVID-19 pneumonia differentiation<sup>5,6</sup>. For the DenseNet-121, however, all reported CCs were very small and were an order smaller than those in VGG-16 and VGG-19 results. These results suggest that SM from DenseNet-121 architecture did not capture meaningful class-specific spatial patterns.

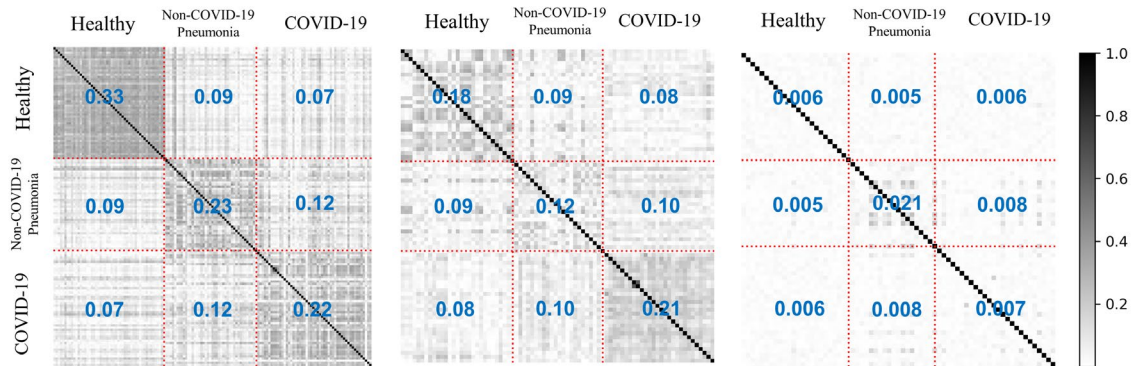


Figure 7. The SM cross-correlation matrix of radiomics-boosted model on test set for left: VGG-16; middle: VGG-19; right: DenseNet121 architectures. The x and y axis represents the sample ID in the test set, sorting with the order of healthy/non-COVID-19 pneumonia/COVID-19 cohorts.

### 3.2 Glioblastoma Post-Resection Survival Prediction

Table 2 summarizes the survival prediction results from vanilla deep learning results, i.e., classification results from deep neural networks' output in Figure 4. In general, when using single MR modality only, deep neural network results in the test set are not promising and the average accuracy result is around 0.5. The best single-input result using 3D FLAIR volumes achieved an accuracy of 0.570 in the test set, but the accuracy in the training set was not satisfying (= 0.54). As a reference, when multiple MR modalities were used as deep neural network input, the combination of FLAIR, T1ce, and T2 were found with the best and yet unsatisfying performances in both 2D and 3D execution. In sum, the current results suggest that GBM survival group prediction based

on classic MR-based deep learning implementation is limited; this observation is consistent with the challenging GBM survival group prediction task in clinic.

**Table 2. Survival group prediction results of deep neural network output as in Figure 4.**

Input	2D		3D	
	Train Accuracy	Test Accuracy	Train Accuracy	Test Accuracy
Flair	0.859	0.509	0.540	0.570
T1	0.772	0.502	0.565	0.483
T1ce	0.897	0.545	0.513	0.568
T2	0.913	0.551	0.604	0.436
FLAIR+T1ce+T2	0.855	0.530	0.521	0.542

The results of SVM execution in Figure 3 using single branch, i.e., radiomics feature ( $RF$ ), deep feature ( $DF$ ), or PSCF (age + tumor region volumes), are summarized in Table 3. When deep features were used, the best accuracy results in the test set were 0.660 in 2D scenario ( $DF_{FLAIR} + DF_{T1ce} + DF_{T2}$ ) and 0.596 in 3D scenario ( $DF_{FLAIR} + DF_{T1} + DF_{T2}$ ); these results are improved from Table 2 results. Note that 2D scenario result is better than 3D scenario result. When radiomics features were used, features selected by dissimilarity analysis ( $RF_{DA}$ ) outperformed the features selected by PCA ( $RF_{PCA}$ ) in both 2D and 3D scenarios. When PSCF was used, the prediction accuracy in test accuracy was 0.638, which is better than results from radiomics feature utilization.

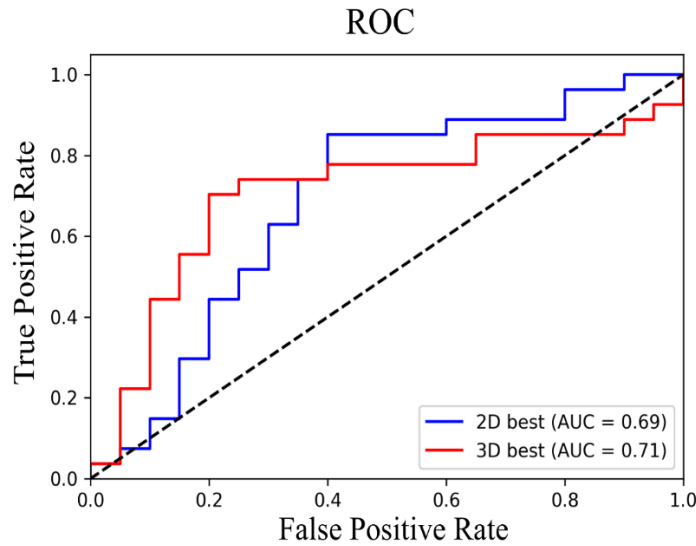
**Table 3. Survival group prediction results of SVM implementation with single data source branch in Figure 3. DF: deep feature; RF: radiomics feature; TS: tumor size**

Input	2D		3D	
	Train Accuracy	Test Accuracy	Train Accuracy	Test Accuracy
$DF_{FLAIR} + DF_{T1(ce)} + DF_{T2}$	0.984	0.660	0.718	0.596
$RF_{PCA}$	0.622	0.511	0.622	0.511
$RF_{DA}$	0.601	0.574	0.590	0.596
$PSCF$	0.697	0.638	0.697	0.638

Table 4 summarizes the key results of the proposed deep learning model in Figure 3 with multiple data branch sources. When radiomics features were utilized together with PSCF, the test group accuracy result was improved from 0.638 (Table 3) to 0.681 in 3D scenario. Similarly, when deep features were utilized together with PSCF, the accuracy was further improved to 0.723. When deep features were included together with PSCF and radiomics features, the achieved test group accuracy was further improved to 0.745 in both 2D scenario ( $PSCF + RF_{DA} + DF_{T1ce}$ ) and 3D scenario ( $PSCF + RF_{DA} + DF_{FLAIR} + DF_{T1ce} + DF_{T2}$ ). The corresponding ROC curves of the identified two models are illustrated in Figure 8. Both curves exhibited similar shapes, and the area-under-curves (AUC) values were 0.69 for the 2D model curve (blue) and 0.71 for the 3D model curve (red). In a nutshell, both 2D and 3D implementation of the proposed multi-dimensional deep learning model achieved the same level of accuracy in GBM survival group prediction.

**Table 4. Survival group prediction results of SVM implementation with multiple data source branches in Figure 3. DF: deep feature; RF: radiomics feature; TS: tumor size. The best results in 2D and 3D scenarios are highlighted.**

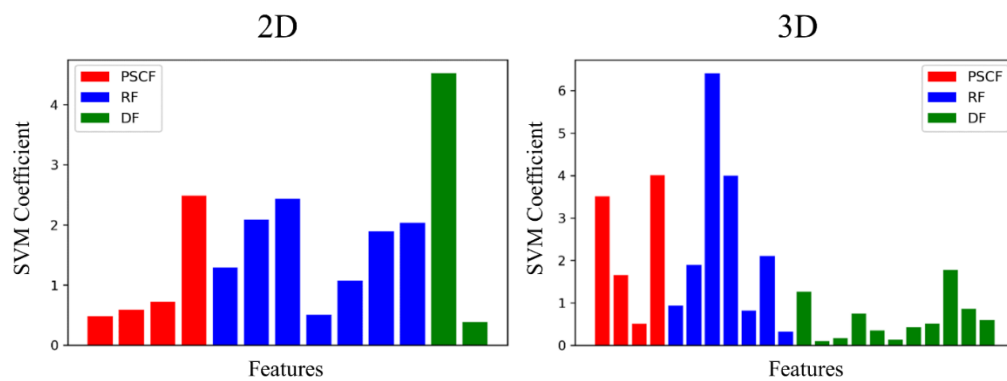
Input	2D		3D	
	Train Accuracy	Test Accuracy	Train Accuracy	Test Accuracy
$PSCF + RF_{DA}$	0.622	0.511	0.718	0.681
$PSCF + DF_{FLAIR}$	0.697	0.638	0.718	0.723
$PSCF + RF_{DA} + DF_{T1ce}$	<b>0.941</b>	<b>0.745</b>	0.707	0.702
$PSCF + RF_{DA} + DF_{T2}$	0.957	0.638	0.755	0.617
$PSCF + RF_{DA} + DF_{FLAIR} + DF_{T1ce} + DF_{T2}$	0.984	0.574	<b>0.777</b>	<b>0.745</b>



**Figure 8. The ROC curves from the identified multi-dimensional deep learning model with the best accuracy results in 2D (blue) and 3D (red) scenarios.**

To evaluate the relative contribution of different features to the final prediction results in SVM execution, we calculated the SVM coefficient of each feature from the aforementioned best models in 2D and 3D scenarios. These results are shown in Figure 9. In general, all three categories' features demonstrated certain contributions to the final prediction results. Although some contribution differences from three categories are

observed, none of the categories made dominant contributions in both 2D and 3D scenarios. The feature with the highest coefficient was found within deep feature group in 2D scenario and within radiomics feature group in 3D scenario, respectively. These results suggest that the proposed model utilized the information from different data source branches in a complementary way without unbalanced focus.



**Figure 9. SVM coefficients of features in 2D (left) and 3D (right) models with the best prediction accuracy results. PSCF: patient-specific clinical feature (age and tumor size); RF: radiomics feature; DF: deep feature**

## **4. Conclusion**

### ***4.1 COVID-19 and non-COVID-19 pneumonia classification***

To our best knowledge, this work is the first of its kind for combining radiomic analysis and deep neural network implementation. The results of this work demonstrated that the inclusion of radiomic feature maps, as a new form of handcrafted imaging biomarker rendering, can improve deep learning-based COVID-19 detection. With the aid of RFMs, we achieve higher model performances for COVID-19/non-COVID-19 pneumonia/Healthy classification with a smaller (812 patients in total) than reported work. For example, Zhang et al.<sup>5</sup> achieved a sensitivity of 88% and a specificity of 79% in COVID-19/non-COVID-19 pneumonia diagnosis with a dataset of 2060 patients. Nishio et al.<sup>51</sup> achieved an accuracy of 83.7% for 3 categories of classification (Healthy/non-COVID-19 pneumonia/COVID-19) using the VGG-16 model trained on 1248 images. Tulin et al.<sup>53</sup> achieved an accuracy of 87.0% for tri-class classification (Healthy/non-COVID-19 pneumonia/COVID-19) using the Darknet-19. In this work, we studied a total of 812 chest X-ray images from 3 public datasets<sup>66-68</sup>, which were not curated by the typical medical image study protocols. As a result, proper image processing is necessary for streamlined deep learning implementation. In particular, we resized all images to 256x256 grid size and normalized all images to 256 gray levels as uint8 format. These operations are standard in digital image processing which will facilitate the data reproducibility of this work. In addition, the robustness of the

developed model was systematically analyzed. For each model design, we trained 50 runs of models using randomly selected training and validation samples following a ratio of 7:1. The small standard deviation ( $< 0.06$ ) of selected metrics and ROC results revealed the enhanced robustness of the developed model, which further demonstrates the potential of the radiomics-boosted deep learning design in clinical situations using different X-ray image data sources.

The deep learning implementations in this work adopted three commonly used deep neural network architectures based on a transfer learning scheme. VGG-16 was first selected for the following reason: 1) it has been widely studied for medical image analysis tasks as a transfer learning scheme; 2) in comparison with other prevalent candidates, VGG-16 possesses a smaller number of trainable parameters under the transfer learning scheme and thus leads to reduced calculation workload for network training; and 3) previous studies<sup>51</sup> reported that VGG-16 achieved the highest accuracy in COVID-19 diagnosis tasks in comparison with several other pre-trained deep neural network architectures. In addition to VGG-16, we studied VGG-19 and DenseNet-121 for the proposed radiomics-boosted design. As reported in Figure 6 and Figure 7, results of VGG-19 after the radiomics-boosted design were similar to the ones in VGG-16 implementation, which can be attributed to the high similarity of network architecture shown in Figure 1<sup>69</sup>. On the other hand, the performance improvement after the radiomics-boosted design was higher in VGG-16 application than in VGG-19 application:

this can be explained by the fact that classification performance from VGG-19 using X-ray image only was higher than ones from VGG-16. Although the total trainable parameter numbers under the transfer learning scheme were approximately the same in VGG-16 and VGG-19, VGG-19 has a larger dimension with more total parameters due to the three additional convolutional layers. Overall, the proposed radiomics-boosted design still improved VGG-19 performance in terms of the 4 classification evaluators as well as model robustness. In the DenseNet-121 study, the improvements after radiomics-boosted design were rather limited. While radiomics-boosted design improved healthy individual classification results with statistical significance, ROC results in non-COVID-19 pneumonia classification showed a slight performance decrease. A plausible explanation is a fact that DenseNet-121 has a larger dimension with more trainable parameters (>70m) than VGG-16 and VGG-19 (~40m) under the transfer learning scheme. In addition, DenseNet-121 has more parameters in total (>75m) than VGG-19 (~55m). As such, a larger X-ray image set might be necessary to exploit the full potential of the proposed radiomics-boosted design in the DenseNet-121 application. Nevertheless, the current results, especially the prominent performance improvement in healthy individual classification, are sufficient to support the benefit of radiomics-boosted design in the DenseNet-121 application.

The inclusion of radiomic feature maps (RFMs) is a key technical innovation. Instead of calculating radiomic features as scalar values from selected volumes in image

space, RFMs capture the anatomy-driven subtle texture variations within ROIs. It has been demonstrated that radiomics are associated with pulmonary function<sup>86</sup> and lung ventilation measurements<sup>72</sup>; as such, the potential functional information in RFMs contributes to the enhanced COVID-19 diagnosis accuracy. The selection of two RFMs from 37 RFMs is more than a trivial task: while direct comparisons of all possible RFM combinations are feasible, it requires a high computational cost without the potential of transferring this technique to other clinical applications. Driven by the hypothesis that certain RFMs can be related to neural network hyperparameters, we selected RFMs based on similarity metrics between RFMs and neural network saliency maps (SMs) which measure the attention pattern of the network implementation. Such a pattern can be used as an auxiliary tool for radiologists in image reading, i.e., highlighting specific regions as visual clues for human reading. This potential human-aid tool can be important in accurate COVID-19/Non-COVID-19 pneumonia differentiation, which can be a challenging task for radiologists using chest X-ray images without volumetric information<sup>5</sup>. RFMs with higher similarities to SMs could emphasize regional information to enhance neural network attention, which increases synchronously with SM results. It is worth mentioning that both VGG-16 and VGG-19 results identified GLCOM Entropy based on SM results, which emphasizes the similarity of the two deep learning architectures. In addition to the current design, it would be of interest to investigate other RFM selection mechanisms that can be complementary to the SM

results (i.e., based on ‘dissimilarity’) for potential performance improvement. DenseNet-121 SM results in Figure 7, however, did not show recognized class-specific SM spatial patterns in comparison with VGG-16 and VGG-19. The absence of such patterns may stem from the potentially limited sample size due to large variable numbers of DenseNet-121. Additionally, ImageNet-based transfer learning may affect SM calculation results as well. Future works of DenseNet-121 analysis, based on full training from scratch or transfer learning from an X-ray-specific dataset, are of our interest to continue SM-based deep learning interpretability studies.

The presented design of combining radiomics analysis and deep neural network implementation may create a new paradigm of the CAD system. The implemented RFM calculation workflow may also enhance neural network performance in other tasks, particularly those where multi-channel imaging data are required as input<sup>87</sup>. Additionally, the proposed method provides a radiomics perspective of deep learning interpretability. The hyperparameters in the neural network are trained without explicit human knowledge intervention and thus are hard to interpret by empirical knowledge. For deep learning-based CAD systems, the ‘black box’ nature impaired the clinical deployments of such systems without clinicians’ confidence. As a step towards deep learning interpretability, we investigated neural network attention information using a radiomics-based analysis. Radiomics has been widely studied as computational imaging biomarkers for disease detection and outcome monitoring<sup>88</sup>, and it has been

demonstrated that radiomic feature spaces can be mathematically decomposed to provide interpretation<sup>89</sup>. Following the saliency map analysis approach in this work, additional parameters can be used to enhance deep learning interpretability, such as histology samples images from biopsy and anatomy contours from radiation therapy. These directions will be studied in future works when appropriate datasets become available.

## ***4.2 Glioblastoma post-resection survival prediction***

In this approach, we developed a novel deep learning model that enables radiomics analysis integration for treatment outcome prediction. To our best knowledge, this work is the first of its kind in multi-dimensional feature fusion for GBM survival group prediction. As handcrafted computational features with known expression, radiomics features reflect quantitative evaluations of high-order statistics that describe image intensity distribution patterns. On the other hand, deep features embedded in a deep neural network learnt image-specific patterns in high-dimensional feature space automatically without analytical expression. The two feature types were hypothesized to possess complementary information due to different feature extraction processes, and thus the combination of the two types in a single machine learning model can potentially improve MR-based GBM survival group prediction with their different extraction. In current clinical practice, basic patient information, including patient age at surgery, tumor location, and tumor size, are reasonable indicators of GBM post-surgery

survival time<sup>90-94</sup>. Our results in Table 3 reported that when using age + tumor region volumes only, the survival group prediction accuracy reached a decent level of 0.638. Results showed that the joint use of radiomics features, deep features, and patient-specific clinical features further improved prediction accuracy to 0.745, suggesting that the great potential of the current study design in clinical practice. It is worth mentioning that the best models in 2D and 3D scenarios achieved the same accuracy of 0.745 in Table 4 and very close ROC AUC results in Figure 8 (2D: 0.69; 3D: 0.71). This result suggests that 2D-based model execution, though with different feature selection results, could be equivalent to 3D-based model execution in terms of results. Intuitively, 3D-based model design is believed to be superior to 2D-based model design with additional volumetric information. With additional high-throughput computation capability, deep neural network may extract sufficient radiography information based on the central 2D axial slice, particularly for GBM patients with large tumor sizes and discernable radiograph patterns. Thus, the 2D execution of the proposed deep learning model could reach the same performance as in 3D execution. For potential clinical application, 2D execution of the current model design might be favored with its lighter model design and less computational load, which are more important when working with larger patient datasets.

Currently, we studied the GBM survival prediction as a bi-class classification problem with 1-yr as short-term/long-term group threshold. This threshold was set based on two reasons: first, it generally agrees with the clinical observation that many patients demonstrated progression after 1-yr survival<sup>58</sup>; and second, it ensures approximately balanced group sizes for radiomics analysis and deep learning implementation. It would be appealing to study survival group prediction as a regression problem to indicate results in terms of days/months. Due to the limited sample size, however, the regression problem design could not converge to extract discernable patterns other than population average results. In future works with enlarged dataset either from a single institution or from multi-institution collaborations, we plan to extend the current feature fusion architecture in regression problem settings.

In radiomics analysis as presented in Figure 3, two different dimension reduction approaches were investigated to condense radiomic information. Results in Table 3 showed that results by dissimilarity analysis (DA) outperformed principal component analysis (PCA) in both 2D and 3D scenarios. The adopted dissimilarity metric intends to select a group of features with less common information to emphasize their complementary roles in modelling. Meanwhile, the extracted principal components in PCA no longer preserve the original form of radiomics features, which may degenerate the embedded analytical information of radiomics features during SVM execution. Thus, dissimilarity analysis may be a better candidate of dimension reduction technique for

radiomics features. During deep feature dimension reduction, dissimilarity analysis was not studied because deep features did not demonstrate meaningful numerical values of dissimilarity, and thus feature selection was not successful; this can be attributed to the fact that deep features are high dimensional statistics without analytical expression, and thus the value-based dissimilarity as 1<sup>st</sup> order statistics is too simple to summarize deep feature patterns. Additionally, deep features are supposed to capture latent differences between the two patient groups during the training, and the dissimilarity measurement may redundantly emphasize the data inhomogeneity. Nevertheless, it might be too immature to rule out the data sample size issue here. When an enlarged dataset becomes available, deep feature dimension reduction will be scrutinized as additional comparison studies.

Besides GBM survival group prediction, the current deep learning model design can be generalized to other oncology applications of outcome prediction based on the technical premise of radiomics analysis and deep learning applicability. In addition, more data source branches, such as genetic profile (i.e., radiogenomic integration<sup>95</sup>), can be added to the current design in Figure 3 as a 'Multi-omics' solution. We anticipate such works may happen soon when relevant data can be available as in potential multi-institutional collaboration.

## References

1. Huang W, Pach D, Napadow V, et al. Characterizing Acupuncture Stimuli Using Brain Imaging with fMRI - A Systematic Review and Meta-Analysis of the Literature. *PLOS ONE*. 2012;7(4):e32960.
2. Buckler AJ, Bresolin L, Dunnick NR, Sullivan DC, Group. A collaborative enterprise for multi-stakeholder participation in the advancement of quantitative imaging. *Radiology*. 2011;258(3):906-914.
3. Nahar J, Imam T, Tickle KS, Ali AS, Chen Y-PP. Computational intelligence for microarray data and biomedical image analysis for the early diagnosis of breast cancer. *Expert Systems with Applications*. 2012;39(16):12371-12377.
4. Cruz-Roa A, Caicedo JC, González FA. Visual pattern mining in histology image collections using bag of features. *Artificial intelligence in medicine*. 2011;52(2):91-106.
5. Zhang R, Tie X, Qi Z, et al. Diagnosis of Coronavirus Disease 2019 Pneumonia by Using Chest Radiography: Value of Artificial Intelligence [published online ahead of print 2020/09/25]. *Radiology*. 2021;298(2):E88-E97.
6. Bai HX, Hsieh B, Xiong Z, et al. Performance of Radiologists in Differentiating COVID-19 from Non-COVID-19 Viral Pneumonia at Chest CT [published online ahead of print 2020/03/11]. *Radiology*. 2020;296(2):E46-E54.
7. Bley TA, Baumann T, Saueressig U, et al. Comparison of radiologist and CAD performance in the detection of CT-confirmed subtle pulmonary nodules on digital chest radiographs. *Investigative radiology*. 2008;43(6):343-348.
8. Jorritsma W, Cnossen F, van Ooijen PM. Improving the radiologist-CAD interaction: designing for appropriate trust. *Clinical radiology*. 2015;70(2):115-122.
9. Bai HX, Wang R, Xiong Z, et al. Artificial Intelligence Augmentation of Radiologist Performance in Distinguishing COVID-19 from Pneumonia of Other Origin at Chest CT [published online ahead of print 2020/04/28]. *Radiology*. 2020;296(3):E156-E165.

10. Roos JE, Paik D, Olsen D, et al. Computer-aided detection (CAD) of lung nodules in CT scans: radiologist performance and reading time with incremental CAD assistance. *European radiology*. 2010;20(3):549-557.
11. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *European journal of cancer*. 2012;48(4):441-446.
12. Chicklore S, Goh V, Siddique M, Roy A, Marsden PK, Cook GJ. Quantifying tumour heterogeneity in 18F-FDG PET/CT imaging by texture analysis. *European journal of nuclear medicine and molecular imaging*. 2013;40(1):133-140.
13. Buvat I, Orlhac F, Soussan M. Tumor texture analysis in PET: where do we stand? *Journal of Nuclear Medicine*. 2015;56(11):1642-1644.
14. Kuhl FP, Giardina CR. Elliptic Fourier features of a closed contour. *Computer graphics and image processing*. 1982;18(3):236-258.
15. Nyúl LG, Udupa JK. On standardizing the MR image intensity scale. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*. 1999;42(6):1072-1081.
16. Haralick; RM, Shanmugam; K, Dinstein; IH. Textural Feature for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*. 1973;SMC-3(6):610-621.
17. Galloway MM. Textural analysis using gray level run lengths. *Computer Graphics and Imaging Processing*. 1975;4(2):172-179.
18. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature communications*. 2014;5(1):1-9.
19. King AD, Chow K-K, Yu K-H, et al. Head and neck squamous cell carcinoma: diagnostic performance of diffusion-weighted MR imaging for the prediction of treatment response. *Radiology*. 2013;266(2):531-538.
20. Zhang H, Tan S, Chen W, et al. Modeling pathologic response of esophageal cancer to chemoradiation therapy using spatial-temporal 18F-FDG PET features,

- clinical parameters, and demographics. *International Journal of Radiation Oncology\* Biology\* Physics*. 2014;88(1):195-203.
21. Coroller TP, Grossmann P, Hou Y, et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. *Radiotherapy and Oncology*. 2015;114(3):345-350.
  22. Lerski RA, Straughan K, Schad L, Boyce D, Blüml S, Zuna I. VIII. MR image texture analysis—an approach to tissue characterization. *Magnetic resonance imaging*. 1993;11(6):873-887.
  23. Petkovska I, Shah SK, McNitt-Gray MF, et al. Pulmonary nodule characterization: a comparison of conventional with quantitative and visual semi-quantitative analyses using contrast enhancement maps. *European journal of radiology*. 2006;59(2):244-252.
  24. Xu R, Kido S, Suga K, et al. Texture analysis on 18F-FDG PET/CT images to differentiate malignant and benign bone and soft-tissue lesions. *Annals of nuclear medicine*. 2014;28(9):926-935.
  25. Fukushima K, Miyake S. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In: *Competition and cooperation in neural nets*. Springer; 1982:267-285.
  26. Lo S-C, Lou S-L, Lin J-S, Freedman MT, Chien MV, Mun SK. Artificial convolution neural network techniques and applications for lung nodule detection. *IEEE transactions on medical imaging*. 1995;14(4):711-718.
  27. Hosseini-Asl E, Keynton R, El-Baz A. Alzheimer's disease diagnostics by adaptation of 3D convolutional network. Paper presented at: 2016 IEEE international conference on image processing (ICIP)2016.
  28. Liu X, Hou F, Qin H, Hao A. Multi-view multi-scale CNNs for lung nodule type classification from CT images. *Pattern Recognition*. 2018;77:262-275.
  29. Khan S, Islam N, Jan Z, Ud Din I, Rodrigues JJPC. A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. *Pattern Recognition Letters*. 2019;125:1-6.

30. Li X, Chen H, Qi X, Dou Q, Fu CW, Heng PA. H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation From CT Volumes. *IEEE Transactions on Medical Imaging*. 2018;37(12):2663-2674.
31. Wang C, Tyagi N, Rimner A, et al. Segmenting lung tumors on longitudinal imaging studies via a patient-specific adaptive convolutional neural network. *Radiotherapy and Oncology*. 2019;131:101-107.
32. Galavis PE, Hollensen C, Jallow N, Paliwal B, Jeraj R. Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters. *Acta oncologica*. 2010;49(7):1012-1016.
33. Yan J, Chu-Shern JL, Loi HY, et al. Impact of image reconstruction settings on texture features in 18F-FDG PET. *Journal of nuclear medicine*. 2015;56(11):1667-1673.
34. Balagurunathan Y, Gu Y, Wang H, et al. Reproducibility and prognosis of quantitative features extracted from CT images. *Translational oncology*. 2014;7(1):72-87.
35. Ammari S, Sallé de Chou R, Balleyguier C, et al. A Predictive Clinical-Radiomics Nomogram for Survival Prediction of Glioblastoma Using MRI. *Diagnostics*. 2021;11(11):2043.
36. Worldometer. *COVID-19 CORONAVIRUS PANDEMIC*.  
<https://www.worldometers.info/coronavirus/>.
37. Wang W, Xu Y, Gao R, et al. Detection of SARS-CoV-2 in Different Types of Clinical Specimens [published online ahead of print 2020/03/12]. *JAMA*. 2020;323(18):1843-1844.
38. Fang Y, Zhang H, Xie J, et al. Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR [published online ahead of print 2020/02/20]. *Radiology*. 2020;296(2):E115-E117.
39. Ai T, Yang ZL, Hou HY, et al. Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. *Radiology*. 2020;296(2):E32-E40.
40. Yoon SH, Lee KH, Kim JY, et al. Chest Radiographic and CT Findings of the 2019 Novel Coronavirus Disease (COVID-19): Analysis of Nine Patients Treated in

- Korea [published online ahead of print 2020/02/27]. *Korean J Radiol.* 2020;21(4):494-500.
41. Xie X, Zhong Z, Zhao W, Zheng C, Wang F, Liu J. Chest CT for Typical Coronavirus Disease 2019 (COVID-19) Pneumonia: Relationship to Negative RT-PCR Testing [published online ahead of print 2020/02/13]. *Radiology.* 2020;296(2):E41-E45.
  42. Wong HYF, Lam HYS, Fong AHT, et al. Frequency and Distribution of Chest Radiographic Findings in Patients Positive for COVID-19. *Radiology.* 2020;296(2):E72-E78.
  43. Guan W, Ni Z, Hu Y, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med.* 2020;382(18):1708-1720.
  44. Shi HS, Han XY, Jiang NC, et al. Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. *Lancet Infect Dis.* 2020;20(4):425-434.
  45. Wang H, Wang L, Lee EH, et al. Decoding COVID-19 pneumonia: comparison of deep learning and radiomics CT image signatures [published online ahead of print 2020/10/24]. *Eur J Nucl Med Mol Imaging.* 2020. doi: 10.1007/s00259-020-05075-4.
  46. Fang X, Li X, Bian Y, Ji X, Lu J. Radiomics nomogram for the prediction of 2019 novel coronavirus pneumonia caused by SARS-CoV-2 [published online ahead of print 2020/07/06]. *Eur Radiol.* 2020;30(12):6888-6901.
  47. Santone A, Belfiore MP, Mercaldo F, Varriano G, Brunese L. On the Adoption of Radiomics and Formal Methods for COVID-19 Coronavirus Diagnosis [published online ahead of print 2021/03/07]. *Diagnostics (Basel).* 2021;11(2).
  48. Wang L, Kelly B, Lee EH, et al. Multi-classifier-based identification of COVID-19 from chest computed tomography using generalizable and interpretable radiomics features [published online ahead of print 2021/01/27]. *Eur J Radiol.* 2021;136:109552.
  49. Minaee S, Kafieh R, Sonka M, Yazdani S, Jamalipour Soufi G. Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning [published online ahead of print 2020/08/12]. *Med Image Anal.* 2020;65:101794.

50. Khan AI, Shah JL, Bhat MM. CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images [published online ahead of print 2020/06/14]. *Comput Methods Programs Biomed.* 2020;196:105581.
51. Nishio M, Noguchi S, Matsuo H, Murakami T. Automatic classification between COVID-19 pneumonia, non-COVID-19 pneumonia, and the healthy on chest X-ray image: combination of data augmentation methods [published online ahead of print 2020/10/18]. *Sci Rep.* 2020;10(1):17532.
52. Amyar A, Modzelewski R, Li H, Ruan S. Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: Classification and segmentation [published online ahead of print 2020/10/17]. *Comput Biol Med.* 2020;126:104037.
53. Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Rajendra Acharya U. Automated detection of COVID-19 cases using deep neural networks with X-ray images [published online ahead of print 2020/06/23]. *Comput Biol Med.* 2020;121:103792.
54. Thakkar JP, Dolecek TA, Horbinski C, et al. Epidemiologic and molecular prognostic review of glioblastoma [published online ahead of print 2014/07/24]. *Cancer Epidemiol Biomarkers Prev.* 2014;23(10):1985-1996.
55. Ostrom QT, Cioffi G, Waite K, Kruchko C, Barnholtz-Sloan JS. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2014–2018. *Neuro-Oncology.* 2021;23(Supplement\_3):iii1-iii105.
56. Stupp R, Mason WP, van den Bent MJ, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma [published online ahead of print 2005/03/11]. *N Engl J Med.* 2005;352(10):987-996.
57. Guillamo JS, Monjour A, Taillandier L, et al. Brainstem gliomas in adults: prognostic factors and classification [published online ahead of print 2001/11/10]. *Brain.* 2001;124(Pt 12):2528-2539.
58. Gutman DA, Cooper LAD, Hwang SN, et al. MR Imaging Predictors of Molecular Profile and Survival: Multi-institutional Study of the TCGA Glioblastoma Data Set. *Radiology.* 2013;267(2):560-569.

59. Kickingereder P, Burth S, Wick A, et al. Radiomic Profiling of Glioblastoma: Identifying an Imaging Predictor of Patient Survival with Improved Performance over Established Clinical and Radiologic Risk Models. *Radiology*. 2016;280(3):880-889.
60. Bae S, Choi YS, Ahn SS, et al. Radiomic MRI Phenotyping of Glioblastoma: Improving Survival Prediction. *Radiology*. 2018;289(3):797-806.
61. Suter Y, Knecht U, Alão M, et al. Radiomics for glioblastoma survival analysis in pre-operative MRI: exploring feature robustness, class boundaries, and machine learning techniques. *Cancer Imaging*. 2020;20(1):55.
62. Cho H-h, Lee HY, Kim E, et al. Radiomics-guided deep neural networks stratify lung adenocarcinoma prognosis from CT scans. *Communications Biology*. 2021;4(1):1286.
63. Chato L, Latifi S. Machine Learning and Deep Learning Techniques to Predict Overall Survival of Brain Tumor Patients using MRI Images. Paper presented at: 2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE); 23-25 Oct. 2017, 2017.
64. Nie D, Lu J, Zhang H, et al. Multi-Channel 3D Deep Feature Learning for Survival Time Prediction of Brain Tumor Patients Using Multi-Modal Neuroimages [published online ahead of print 2019/02/02]. *Sci Rep*. 2019;9(1):1103.
65. Tang Z, Xu Y, Jin L, et al. Deep Learning of Imaging Phenotype and Genotype for Predicting Overall Survival Time of Glioblastoma Patients [published online ahead of print 2020/01/07]. *IEEE Trans Med Imaging*. 2020;39(6):2100-2109.
66. Google. *COVID-19 X-ray dataset*.  
<https://drive.google.com/file/d/1AfC8emc3mGCfIYo1jK-R6suK7a2Kv2m2>.
67. Cohen J, Morrison P, Dao L. COVID-19 Image Data Collection. *arXiv*. 2020. doi: 2003.11597.
68. Paul M. *Chest X-Ray Images (Pneumonia)*.  
<https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>.
69. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:14091556*. 2014.

70. Huang; G, Liu; Z, Maaten; Lvd, Weinberger KQ. Densely Connected Convolutional Networks. *arXiv*. 2018.
71. Deng J, Dong W, Socher R, et al. ImageNet: A Large-Scale Hierarchical Image Database. In: *Cvpr: 2009 Ieee Conference on Computer Vision and Pattern Recognition, Vols 1-4*. doi: 10.1109/cvpr.2009.5206848 New York: Ieee; 2009:248-255.
72. Yang Z, Lafata K, Chen X, et al. Quantification of lung function on CT images based on pulmonary radiomic filtering. *arXiv*. 2021. doi: arXiv:2105.11171.
73. Chang Y, Lafata K, Wang C, et al. Digital phantoms for characterizing inconsistencies among radiomics extraction toolboxes. *Biomedical Physics & Engineering Express*. 2020. doi: 10.1088/2057-1976/ab779c.
74. Zwanenburg A, Vallières M, Abdalah MA, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology*. 2020;295(2):328-338.
75. Karen S, Andrea V, Andrew Z. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. *arXiv*. 2014. doi: arXiv:1312.6034v2.
76. Menze BH, Jakab A, Bauer S, et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE transactions on medical imaging*. 2014;34(10):1993-2024.
77. Bakas S, Akbari H, Sotiras A, et al. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific data*. 2017;4(1):1-13.
78. Yang Z, Lafata K, Chen X, et al. Quantification of lung function on CT images based on pulmonary radiomic filtering. *arXiv preprint arXiv:210511171*. 2021.
79. Berenguer R, Pastor-Juan MDR, Canales-Vázquez J, et al. Radiomics of CT features may be nonreproducible and redundant: influence of CT acquisition parameters. *Radiology*. 2018;288(2):407-415.
80. Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*. 2015;115(3):211-252.

81. Pan SJ, Yang Q. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*. 2009;22(10):1345-1359.
82. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. Paper presented at: International Conference on Medical image computing and computer-assisted intervention2015.
83. Zhou Z, Sodha V, Rahman Siddiquee MM, et al. Models Genesis: Generic Autodidactic Models for 3D Medical Image Analysis. Paper presented at: Medical Image Computing and Computer Assisted Intervention – MICCAI 2019; 2019//, 2019; Cham.
84. Zhou Z, Sodha V, Pang J, Gotway MB, Liang J. Models Genesis. *Medical Image Analysis*. 2021;67:101840.
85. Duan K, Keerthi SS, Poo AN. Evaluation of simple performance measures for tuning SVM hyperparameters. *Neurocomputing*. 2003;51:41-59.
86. Lafata KJ, Zhou Z, Liu JG, Hong J, Kelsey CR, Yin FF. An Exploratory Radiomics Approach to Quantifying Pulmonary Function in CT Images [published online ahead of print 2019/08/10]. *Sci Rep*. 2019;9(1):11509.
87. Wang C, Liu C, Chang Y, et al. Dose-Distribution-Driven PET Image-Based Outcome Prediction (DDD-PIOP): A Deep Learning Study for Oropharyngeal Cancer IMRT Application. *Frontiers in Oncology*. 2020;10(1592).
88. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine [published online ahead of print 2017/10/05]. *Nat Rev Clin Oncol*. 2017;14(12):749-762.
89. Lafata KJ, Chang Y, Wang C, et al. Intrinsic radiomic expression patterns after 20 Gy demonstrate early metabolic response of oropharyngeal cancers [published online ahead of print 2021/05/08]. *Med Phys*. 2021;48(7):3767-3777.
90. Lacroix M, Abi-Said D, Fournay DR, et al. A multivariate analysis of 416 patients with glioblastoma multiforme: prognosis, extent of resection, and survival. *Journal of neurosurgery*. 2001;95(2):190-198.
91. Gittleman H, Lim D, Kattan MW, et al. An independently validated nomogram for individualized estimation of survival among patients with newly diagnosed

glioblastoma: NRG Oncology RTOG 0525 and 0825. *Neuro-oncology*. 2017;19(5):669-677.

92. Batchelor TT, Betensky RA, Esposito JM, et al. Age-Dependent Prognostic Effects of Genetic Alterations in Glioblastoma. *Clinical Cancer Research*. 2004;10(1):228-233.
93. Walid MS. Prognostic factors for long-term survival after glioblastoma. *The Permanente Journal*. 2008;12(4):45.
94. Awad A-W, Karsy M, Sanai N, et al. Impact of removed tumor volume and location on patient outcome in glioblastoma. *Journal of neuro-oncology*. 2017;135(1):161-171.
95. Lafata KJ, Corradetti MN, Gao J, et al. Radiogenomic Analysis of Locally Advanced Lung Cancer Based on CT Imaging and Intratreatment Changes in Cell-Free DNA. *Radiology: Imaging Cancer*. 2021;3(4):e200157.