IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 17, 2021, accepted May 31, 2021, date of publication June 10, 2021, date of current version June 30, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3088234

Genotype-Guided Radiomics Signatures for Recurrence Prediction of Non-Small Cell Lung Cancer

PANYANAT AONPONG[1], YUTARO IWAMOTO[1], (Member, IEEE), XIAN-HUA HAN[2], (Member, IEEE), LANFEN LIN[3], AND YEN-WEI CHEN[1,4], (Member, IEEE)

[1]Graduate School of Information Science and Engineering, Ritsumeikan University, Kusatsu 525-8577, Japan
[2]Artificial Intelligence Research Center, Yamaguchi University, Yamagushi 753-8511, Japan
[3]College of Computer Science and Technology, Zhejiang University, Hangzhou 310000, China
[4]Research Center for Healthcare Data Science, Zhejiang Laboratory, Hangzhou 310027, China

Corresponding authors: Yen-Wei Chen (chen@is.ritsumei.ac.jp) and Lanfen Lin (llf@zju.edu.cn)

This work was supported in part by the Grant-in Aid for Scientific Research from the Japanese Ministry for Education, Science, Culture and Sports (MEXT) under Grant 20KK0234, Grant 18H03267, and Grant 20K21821.

This work did not involve human subjects or animals in its research.

ABSTRACT Non-small cell lung cancer (NSCLC) is a serious disease and has a high recurrence rate after surgery. Recently, many machine learning methods have been proposed for recurrence prediction. The methods using gene expression data achieve high accuracy rates but expensive. While, the radiomics features using computer tomography (CT) image is a cost-effective method, but their accuracy is not competitive. In this paper, we propose a genotype-guided radiomics method (GGR) for obtaining high prediction accuracy at a low cost. We used a public radiogenomics dataset of NSCLC, which includes CT images and gene expression data. Our proposed method is two steps method that uses two models. The first model is a gene estimation model, which is used to estimate the gene expression from radiomics features and deep features extracted from CT images. The second model is used to predict the recurrence using the estimated gene. The proposed GGR method is designed based on hybrid features which is the fusion of handcrafted- and deep learning-based features. The experiments demonstrated that the prediction accuracy can be improved significantly from 78.61% (existing radiomics method) and 79.09% (ResNet50) to 83.28% by the proposed GGR.

INDEX TERMS Non-small cell lung cancer, prediction of recurrence, radiogenomics, genotype-guided radiomics.

I. INTRODUCTION

Non-small cell lung cancer (NSCLC) is one of the most fatal diseases in humans [1], [2]. The NSCLC occurs in about 80% - 85% of lung cancer patients and can be treated by surgery [1]. There is a high risk of recurrence even after treatment with NSCLS [3]. According to statistics, the recurrence of NSCLC cancer causes the death of about 50% of patients [4]. With an accurate prediction of recurrence, doctors will be able to treat patients on time. If doctors and patients have an accurate prediction of the risk of recurrence, the doctor will be able to prepare and look after patients appropriately [5].

Several machine learning methods have been proposed to develop a computer-aided diagnosis. In 2012, Lambin *et al.* proposed the concept of radiomics [6]. Lambin *et al.* used the medical image to extract high-throughput mining of quantitative image features (sometimes called handcrafted features or radiomics signatures) [6]. Thus, the extracted features were used to support the systems to improve diagnostic and predictive accuracy [6]. Kumar *et al.* surveyed a research to support the performance of the radiomics in several kinds of image and the phenotype including CT image and NSCLC [7].

The associate editor coordinating the review of this manuscript and approving it for publication was Emre Koyuncu.

Kumar *et al.* also found a correlation between the extracted features and genotypes [7]. In 2014, Aerts *et al.* found the association of the handcrafted features and underlying gene expression association [8]. In 2016, Gillies *et al.* re-emphasized the radiomics core ideas and reported the radiomics potential power to facilitate better clinical decision-making in cancer patients [9]. In 2018, Alahmari *et al.* was successful in the use of radiomics to predict lung nodule malignancy [10]. In 2019, Wang *et al.* used the radiomics technique to make the preoperative recurrence prediction of NSCLC using principal component analysis (PCA) [11] for feature selection then used several machine learning methods such as decision tree, random forest, etc. to make the classification [12]. In 2020, Lee *et al.* applied the radiomics technique using the Relief-F method [13] for feature selection and performed classification using machine learning methods [14]. In 2021, Christie *et al.* used the least absolute shrinkage and selection operator (LASSO) [15], [16] for both feature selection and classification [17]. To support and improve the capacity of radiomics signature, many radiomics signature-based methods have been proposed [10], [18]–[20]. All studies demonstrated that the handcrafted feature is a potential biomarker for the recurrence prediction of NSCLC.

Recently, many deep learning-based approaches were also used for malignant tumor classification tasks. In 2014, Karen *et al.* proposed the model named the very deep convolutional networks or known as VGG [21]. The VGG uses the convolutional layers to extract the image deep features and uses the neural networks to classify [21]. In 2016, He *et al.* proposed a well-known deep learning model named deep residual learning networks or known as ResNet [22] (more details in section III (B 2)). In 2018, Gao *et al.* also proposed the model named densely connected convolution networks, known as DenseNet [23]. Many researchers in the medical field also proved that the deep learning models are excellent at working with medical imaging [24]–[26].

In our previous research [27], we extracted features through handcrafted and deep learning modules to use the benefits of both feature types.

However, the previously mentioned techniques use only the image information, the classification performance is limited [28]. Many modern studies attempted to use gene expression instead of using only the image. Lu *et al.* [4] and Lee *et al.* [29] proposed the methods that use gene expression to increase the performance of the models. Subramanian used the fusion of image and gene expression together to maximize the prediction accuracy [3]. Many more genomics-based methods were also proposed to improve the performance of the classification and regression tasks [30]–[32].

Although the genomics-based methods achieved high prediction accuracy, an invasive process is needed to gain gene expression. Furthermore, the cost of the gene expression is high compared to the CT image. In many cases, some required gene expressions are undetected. The gene expression cannot be obtained from the patients who are unable to undergo surgery. In paper [3], they use the combination of genomics and radiomics, which has produced the highest results. The limitations in [3] are the possibility of operation, availability of gene expression data, and the high cost due to the complexity of the examination, this is an invasive diagnostic method [33].

Thus, the primary aim of our study is to propose a technique that can maximize the recurrence prediction accuracy using CT image and gene expression in the training phase and CT image in the testing phase. We use the association of the image and gene expression [7], [8], Due to the strong relationship between the gene expression and the phenotype [3], [4], [29], [31], [32] we have developed a new method. We propose a genotype-guided radiomics-based approach (GGR). Our proposed method consists of two models. The first model is used to estimate the gene expression from CT images. The second model is used to predict the recurrence based on the estimated gene expression which we have achieved from the first model. After the GGR is trained (using image, gene expression, and recurrence label), the GGR does not require the gene expression data in the testing phase. The GGR model significantly improves the accuracy and provides the possibility of examining the patients who are unable to undergo the surgery.

II. MATERIALS

In this research, we used a radiogenomics dataset of NSCLC [34], which is now publicly available for use in The Cancer Imaging Archive (TCIA) [35], [36] for the experiment. The R01 cohort from Stanford University School of Medicine and Palo Alto Veterans Affairs Healthcare System were recruited between April 7th, 2008, and September 15th, 2012. Subjects signed written consent forms according to the guidelines of institutions' IRBs. This dataset was collected from the NSCLC cohort of 162 subjects and the data of every subject can be separated into two types including image data and gene expression data [34].

A. IMAGE DATA AND IMAGE PRE-PROCESSING

Image data is collected from preoperative CT scans at the Stanford University Medical Center. Imaging data has a thickness of 0.625–3 mm and an X-ray tube current at 124–699 mA at 80–140 kVp. The type of CT image is in the DICOM format and consists of 162 patients in total [34]. Some samples were eliminated due to established rules, including the patient without tumor mask and the patient without gene expression data, recurrence data, or died before recurrence occurs due to unknown reasons.

The initial segmentation was received from the axial CT image series using the unpublished automatic segmentation algorithm [34]. All these divisions are seen by thoracic radiologists with experience over 5 years and corrected the data as needed using ePAD software [34]. The final segmentation is revised and made the final approval (tumor discussion and amended as appropriate) by additional thoracic radiologists [34]. The Tumor masks are stored

in DICOM format [34]. Some examples of image data and tumor mask areas are shown in Figure 1.

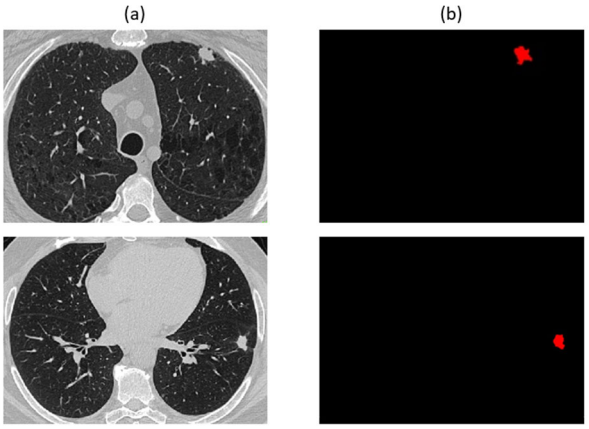

FIGURE 1. Two examples of image data (a) image data and (b) tumor mask area.

For image pre-processing, we select the slice with the largest tumor mask area and its two neighbor slices (the above and below slices) for each volume. These three images are fed into the network as three channels. We use three slices for our research because we want to include 3D information of the volume data. In our previous work [37], we have shown that the performance of multiple slices is better than that of a single slice. Detailed information about feature extraction will be described in Section III A. Here, the Hounsfield unit observation range setup was performed to get the suitable radio information. We cut the intensity information outside the range of -1000 HU $\sim$ $+400$ HU (Hounsfield Unit). This range covers the information that is needed from the CT Image [38], and normalized value in the entire three slices to greyscale value with range 0 to 255 using the linear transformation (Figure 2). The CT image and its corresponding tumor mask image are multiplied. Then we crop the area outside the bounding box around the masked image and resized the cropped image to 224×224 pixels.

B. GENE EXPRESSION DATA AND GENE SELECTION

To examine gene expression, biologists must collect the tumor samples. All tumor samples were collected during surgery from volunteers taking not any medications or drugs. After tumor removal, the surgeon cuts 3-5 mm thick pieces along the longest axis of the cut tissue, which is frozen within 30 minutes after removal. The tissue was analyzed using the expression of RNA sequences.

The gene expression data used in this study is RNA-sequencing data. The gene expression data availability is depending on the availability and quality of existing tissues. The gene expression was sequenced by HiSeq 2500 (Illumina) following the manufacturer's instructions. The 130 tissue-set samples have been sequenced in 3 batch sizes: 16, 66, and 48. The gene expression was pre-processed by Centrillion Bioscience. Finally, the gene expression is

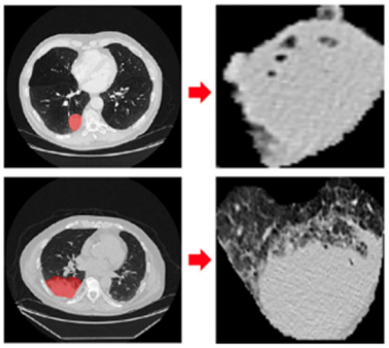

FIGURE 2. Full CT images with the red masking area (left) and the images after pre-processing (crop and normalization; right).

estimated in Fragments per Kilobase of Transcript Per Million (FPKM) unit [34]–[36], [39]. 22,127 genes were provided for each patient [34]. In the dataset, there are only 130 patients from 162 patients whose gene expression data exist [34]. Most patients' gene expression data has shown unclear expression, in some patients, indicated by N/A. These ambiguous gene expressions are removed from our work. Finally, 5,587 gene expression data are available in the study dataset for each patient. After examining the data, the dataset includes images from 88 patients that meet our experimental requirement. The dataset details are presented in Table 1.

TABLE 1. Clinical characteristics of the screened subjects.

		Total n = 88 patients
Age (year)		46~85 (median = 69)
Gender	Male	64 (72.72%)
	Female	24 (27.27%)
Smoking status	Nonsmoker	15 (17.05%)
	Current	14 (15.91%)
	Former	59 (67.05%)
Cell type	ADC	68 (77.27%)
	SQC	17 (19.30%)
	Not specified	3 (3.41%)
Recurrence	No	29 (32.95%)
	Yes	59 (67.05%)
		Total n = 88 patients

Due to the large gene dataset that contains more than 20,000 gene expression data for each patient, the massive gene expression data will significantly increase the computational cost and decreases the prediction accuracy. In other words, spending time for the estimation of 20,000 genes is worthless. Thus, we must select only the associated gene that we want to estimate before training the model.

Based on the large data of genes from a patient, relevant genes need to be selected before declaring the regression model. In other words, we will not declare the regression model for the non-relevant gene. By the gene selection, the genes which are not associated with the recurrence of NSCLC are removed. We use several feature selection

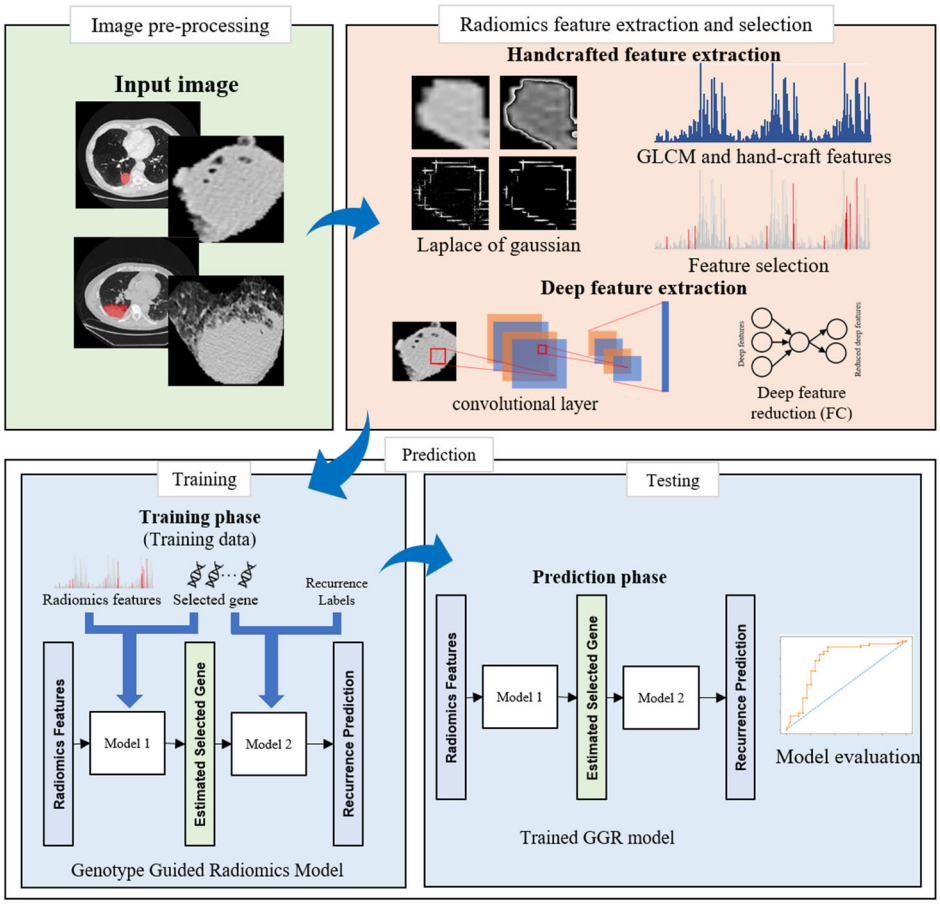

FIGURE 3. **Overview of genotype-guided radiomics (GGR) for recurrence prediction of non-small-cell lung cancer.**

methods to select the associated genes including LASSO [15]–[17], F-test (ANOVA) [40], [41], CHI-2 [42], and the intersection of the three feature selection methods. Finally, we select 74 related genes. The detailed results will be reported in Section IV.

III. METHODS

The workflow of the proposed method is shown in Figure 3. The whole process of the method consists of three parts: (1) data pre-processing; (2) radiomics feature extraction and selection (input features); (3) prediction (gene estimation and recurrence prediction models). Our proposed method is mainly contributed to the third prediction part. We called this proposed prediction part the genotype-guided radiomics (GGR) model. The traditional radiomics-based method uses one model to predict the recurrence from the CT images, the proposed genotype-guided radiomics (GGR) method consists of two models. Both models were trained using the NSCLC public radiogenomic dataset [34], which includes CT images and corresponding gene expression data. The key proposal of the GGR is that we only use CT images to predict recurrence, while the models are trained by pairs of gene expression data and CT images in the training phase. That means the trained GGR model can represent some relationship between the CT image and its gene expression. Therefore, we can use the gene expression estimated from the CT image for recurrence prediction even though we do not have gene expression data in the test phase.

Since the first part (data pre-processing) has been described in Section II, we will focus on the second and the third parts in this section.

A. RADIOMICS FEATURE EXTRACTION AND SELECTION

The image feature extraction is used to reveal the information from the image. We perform two types of feature extraction, including handcrafted feature extraction and deep feature extraction as shown in Figure 4.-6., for the gene estimation model. The handcrafted feature extraction will show the information at the macro-level from the image and the deep feature extraction will show the information at the micro-level [27]. Using both features allows us to get higher gene estimation and recurrence prediction performance.

1) HANDCRAFTED FEATURE EXTRACTION

We currently have 3 slices of cropped CT image. On each slice, 150 radiomics features will be extracted using gray level co-occurrence matrix (GLCM) in 4 directions, 0, 45, 90, and 135 degrees, and histogram-based

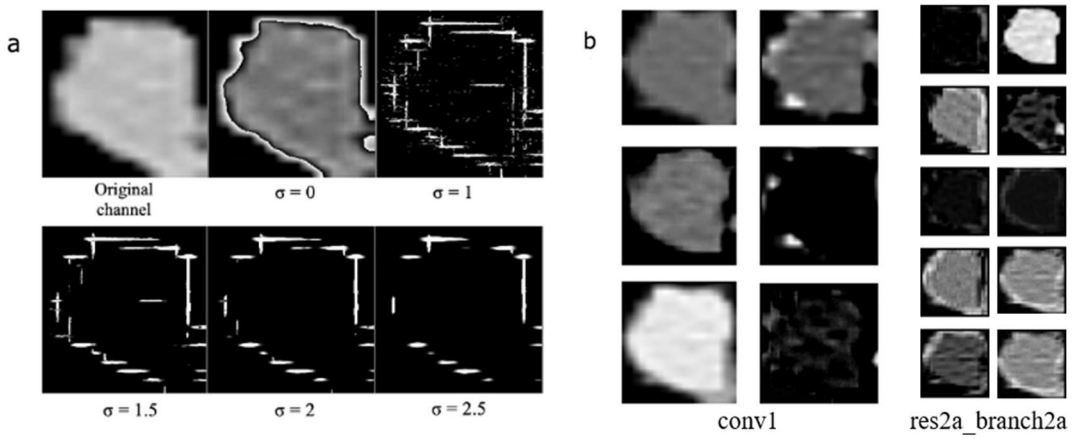

FIGURE 4. (a) The result of five filters LoG on one slice. (b) The examples of convolution products from intermediate layers of deep learning-based features extraction (conv1 and res2a_branch2a layers).

statistics [16], [18], [38], [43]. In the calculation of each feature, Laplace of Gaussian (LoG) has been performed.

The LoG is the combination between the Laplacian filter and the Gaussian filter. Technically, the Gaussian filters make images smoother to reduce noise, and Laplacian filters focus on areas that change quickly or work like edge detection. To calculate LoG, we first calculate the Gaussian distribution using equation (1) [44].

$$G(x, y; \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \tag{1}$$

where x and y are the coordinates on the x and y axis, respectively, and σ is the standard derivation or the biometrics filter parameter [44]. The Gaussian scale-space representation of the image $f(x, y)$ can be computed as equation (2)

$$L(x, y; \sigma) = f(x, y) * G(x, y; \sigma) \tag{2}$$

From equation (2), $L(x, y; \sigma)$ is the Gaussian scale-space of the image $f(x, y)$, and $*$ is a convolutional operator [44].

From both equation (1) and equation (2), we apply the Laplacian operator ∇^2 to the Gaussian scale-space representation of the image. We can calculate ∇^2 as equation (3).

$$\nabla^2 = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \tag{3}$$

Then, the LoG can be summarized as equation (4).

$$\nabla^2 G(x, y) = -\frac{1}{\pi\sigma^4}\left[1 - \frac{x^2 + y^2}{2\sigma^2}\right] \exp^{-\frac{x^2+y^2}{2\sigma^2}} \tag{4}$$

From LoG equations (1) - (4), five different σ filters ($\sigma = 0$, 1, 1.5, 2, 2.5; Figure 4 (a)) were assigned to the calculation.

GLCM features of four directions, which involves angles at 0°, 45°, 90°, and 135°, were calculated from each LoG filter. Radiomics features include contrast, entropy, relationships, homogeneity, and energy are calculated from each degree of angle. In addition, the intensity features use the first-order statistic to quantify the radiomics feature. The first-order statistics include mean, standard derivation (SD), percentiles mean (10, 25, and 50 percentiles), percentiles SD (10, 25, and 50 percentiles), kurtosis, and skewness [16], [19]. The structure of radiomics features is shown in Figure 5. The radiomics features are extracted from the selected three slices. Accordingly, 450 radiomics features are extracted from one patient.

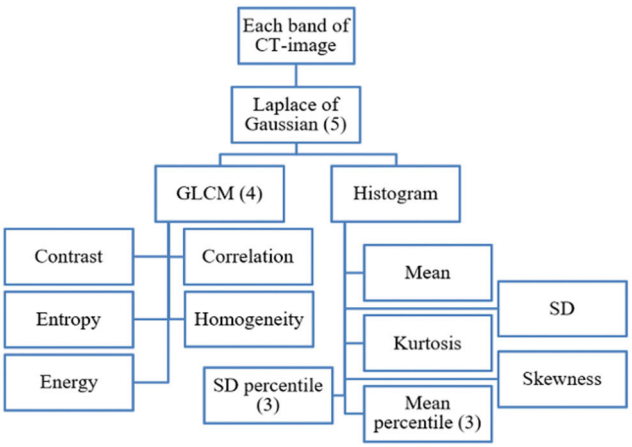

FIGURE 5. The radiomics feature structure from each band of CT image. (The number in parentheses indicates the number of the features' groups.)

After radiomics feature extraction, the extracted radiomics features which is not related to the recurrence are removed. In the conventional radiomics methods including [3], [8], [12]–[19], [20], [27], the feature selection is used to select the radiomics features of the top association to the recurrence of the NSCLC. As reported in [27], we found that the F-test (ANOVA; Analysis of Variance) has shown the highest accuracy and AUC in recurrence prediction using CT image compared to various feature selection methods. In this work, we used F-test [40], [41] to select the associated radiomics features to predict the gene related

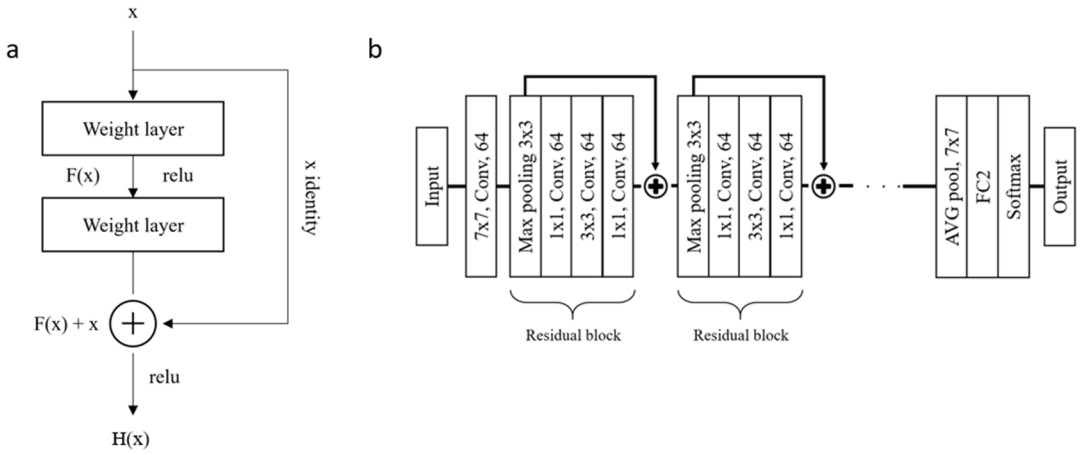

FIGURE 6. (a) A single residual block with the skip connection used in ResNet. (b) The ResNet structure.

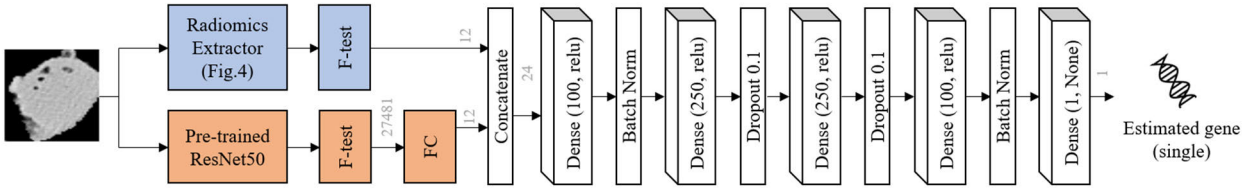

FIGURE 7. The DNN regression model's structure used to estimate a single gene from the extracted and selected features.

to the recurrence. With F-test, we can reduce the number of radiomics features from 450 radiomics features to only 12 associated radiomics features. The selected 12 features are used to estimate the gene expression. The radiomics feature extractions are shown in the left part (blue color) of Figure 7. Our radiomics codes are now released and available at URL: https://github.com/aonpong/radiomics.

2) DEEP FEATURE EXTRACTION

In deep feature extraction, we applied a ResNet50 structure [22]. The network is initialized using the pre-trained weights of ImageNet dataset [45] and eliminated the fully connected layers. Many studies reported the pre-training model using the ImageNet can improve the ability of medical image classification tasks [46], [47].

ResNet50 is a convolutional neural network model that is 50 layers deep [22]. The ResNet50 consists of many residual blocks. In each block, it has convolution modules and a skip connection that pass the information calculated from the previous block to the next block. It has been observed in earlier layers, the learned features correspond to the lower semantic information. Without skip connection, the information will turn to abstract because of the long-chain connection. In other words, the non-skip connection is the cause of very small gradients in the deeper layers and has a probability to become zero. For this reason, we cannot update the early layer at all. As shown in Figure 6 (a), x represents the input, $H(x)$ represents the output, then, the residual knowledge of the shortcut connection is $F(x) = H(x) - x$. The skip connection can reduce loss from the deep calculation, which is affected by the chain rule in the backpropagation because this provides the alternative path for the gradient and allow the information to pass through. The information that passed the skip connection will be attached to the calculated information via addition. The structure of the ResNet50 model is shown in Figure 6 (b).

The NSCLC recurrence-related output features extracted from the ResNet50 will be selected using the F-test method [40], [41] as the feature selection, 27,481 deep features were selected. The features are further reduced to 12 using a fully connected (FC) layer to have the same dimension as the handcrafted features. The deep feature extractions are shown in the left part (orange color) of Figure 7. Finally, the 12 deep features and 12 handcrafted radiomics features are used for gene estimation as shown in Figure 7.

B. PREDICTION (GENE ESTIMATION AND RECURRENCE PREDICTION)

Unlike the traditional radiomics-based method, which uses one model to predict the recurrence from the CT images, the proposed genotype-guided radiomics (GGR) method consists of two models. The first model is used to estimate the gene expression data from CT images and the second model is used to predict the recurrence using the estimated gene.

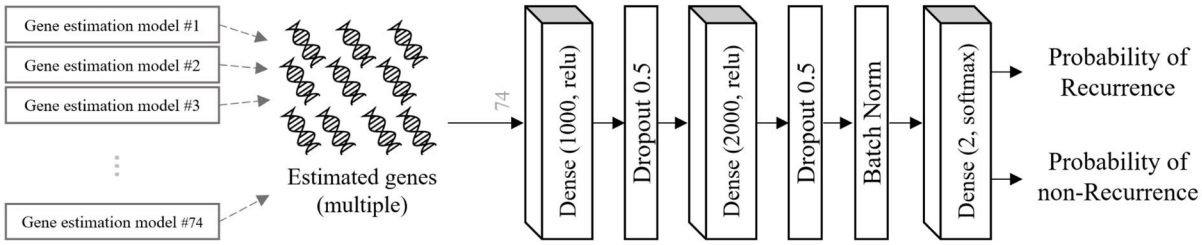

FIGURE 8. The GGR's structure uses to predict the recurrence using estimated genes. The grey blocks show the source of estimated genes (previous gene estimation models). The grey numbers show the number of features that pass the process.

1) GENE ESTIMATION (THE FIRST MODEL)

We use a deep neural network (DNN) regression model to estimate gene expression as shown in Figure 7. As we described in the previous sub-section (Section III A), we use two features (handcrafted and deep features) as the inputs of the model. Both features are 12-dimensional vectors. We concatenate them to form a 24-dimensional feature for gene estimation (regression).

The gene expression data is only required to train the model (the mapping function between the CT image and gene expression) but not for the testing phase because we can estimate the gene from the CT image using the trained model. Note that one regression model is used to estimate one gene. As we described in Section II B, we select 74 related genes to reduce the computation cost and enhance the prediction accuracy. That means we trained 74 regression models for 74 genes (one model for each gene). All regression models have the same structure as shown in Figure 7 but have different weights for the individual gene estimation. For all the DNN regression models, we use the mean square error [48] as a loss function which can be calculated using equation (5).

$$\mathcal{L}_{gene} = \frac{1}{m}\sum_{i=1}^{m}\left\|\mathbf{G}_i - \hat{\mathbf{G}}_i\right\|^2 \tag{5}$$

where $\mathcal{L}_{gene}$ is the mean squared error loss between the actual gene expression ($\mathbf{G}$) and the predicted gene ($\hat{\mathbf{G}}$), m is the number of training samples, $\|.\|$ is the L2 norm. The learning rate was set to a very small value of 5e-6 to fit every gene estimation. The decay and momentum were set to 1e-6 and 0.9, respectively.

2) RECURRENCE PREDICTION

After gene estimation, the estimated gene expression data are used as inputs of the recurrence prediction model (the second model). The recurrence prediction is done as a task of two-class classification (i.e., recurrence and non-recurrence). We proposed an ANN-based model to predict the recurrence of NSCLC using estimated genes as shown in Figure 8. We use stochastic gradient descent (SGD) [49] as an optimizer. The learning rate, decay, and momentum were set to 0.05, 1e-6, and 0.9, respectively. The loss function used in GGR is binary cross-entropy loss [50] which can be calculated by equation (6).

$$\mathcal{L}_{bce} = -\frac{1}{m}\sum_{i=1}^{m}\left[r_i \ln \hat{r}_i + (1 - r_i)\ln(1 - \hat{r}_i)\right] \tag{6}$$

where $\mathcal{L}_{bce}$ is binary cross-entropy loss, m is number of training samples, r is actual recurrence label and $\hat{r}$ is recurrence prediction output.

IV. EXPERIMENTS

Every experiment in this study is performed based on ten-fold cross-validation to find the average performance. On each fold, we used CT images and gene expression data of 79–80 patients as training sets and only the CT image of 7-8 patients as the validation set (gene expression data does not require). We also used an area calculation under the characteristic curve of the receiver operating characteristics (AUC) [51] to assess the efficiency of each candidate model. We performed these experiments on a computer driven by CPU Intel®core^{TM}i7-8700k @3.20-4.60GHz, 48 GB of random-access memory (RAM), and RTX 2060 graphic accelerator. We used Keras-GPU library version 2.2.4 on python 3.6 to perform these experiments. The related work models are compared to the proposed GGR methods using the same control data set.

A. GENE SELECTION RESULTS

We performed this experiment to observe the gene effectiveness and to choose the best gene set, which is the most relevant to the recurrence of NSCLC. We compared four feature selection methods, including non-selected, LASSO, F-test, CHI-2, and the intersection of the three. The gene features which have LASSO's zero coefficient were removed [16], [19]. The P-values from F-test and CHI-2 were set the threshold at P-value < 0.02 [40]–[42] The genes are selected based on the results obtained using the intersection of all the three methods including LASSO, F-test, and CHI-2. We test each output gene obtained from the various feature selection method to directly classify the recurrence. The gene set that provided the best accuracy will be chosen to predict the recurrence in model 2. The results of gene selection performance are presented in Table 2. Since the method of the intersection of the three achieves the best performance, we used 74 genes selected by the intersection

TABLE 2. Performance of the radiomics method using selected real gene data from different gene selection methods.

Feature selection method	Selected genes	Accuracy
Non-selected	5587	81.14%
LASSO	1123	82.97%
F-test	131	86.89%
CHI-2	2325	83.39%
Intersection of LASSO, F-test, and CHI-2 (P-value < 0.02)	**74**	**94.30%**

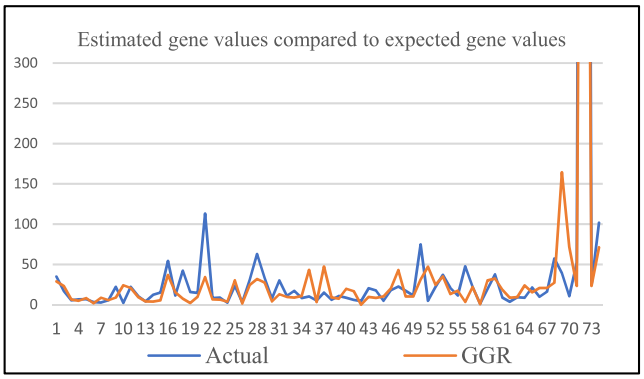

FIGURE 9. Comparison between estimated gene values and actual gene values from one patient as an example.

of the three methods for our studies. The 74 selected genes list has shown in Table 3.

B. GENE ESTIMATION RESULTS

This section shows the comparison between the estimated gene expression data derived from the model and the real gene expression data in FPKM. Due to the complexity of the data, Figure 9. shows the comparison between the estimated genes using the hybrid of handcrafted features and deep features, and the actual values of one sample. In Figure 9, the x-axis shows 74 genes of one sample patient, which are selected by the intersection of the three gene selection methods, and the y-axis shows the actual gene expression FPKM values (the blue line), the estimated gene expression from the fusion of handcrafted features and the deep features in FPKM values (the orange line; average MSE = 29293.07). The graph depicts that the majority of the estimated genes show satisfying estimation performance.

C. ABLATION STUDY

In this section, we validated the effectiveness of each key component for recurrence prediction. The ablation results are summarized in Table 4. Model 1 and model 2 are baseline models with handcrafted features or deep features only, respectively. Model 3 is a model with combined handcrafted and deep features, which is our first contribution but without gene guidance. Model 4 (i.e., the proposed GGR model) is a model with gene guidance, which is our second contribution. As shown in Table 4, we can see that the AUCs of model 1 and model 2 (baseline models) are about 0.66 and 0.67, respectively. The AUC can be improved to 0.71 by combining the handcrafted and deep features (model 3). The AUC can be further improved to 0.77 using gene guidance (model 4).

D. COMPARISON WITH STATE-OF-ART METHODS

We also compared the proposed GGR method with state-of-the-art methods. The comparison results are summarized in Table 5. The first seven methods are handcrafted radiomics-based methods and the 8^{th}-10^{th} methods are deep learning-based methods. The proposed method is indicated in bold. Furthermore, Figure 10 shows the comparison between the average ROC curves of 10-fold cross-validation to the other recurrence prediction methods. As shown in Table 5 and Figures 10., deep learning-based methods [25], [27] achieved better performance than handcrafted radiomics-based methods [12], [14], [17], [27] The proposed GGR method improves the performance significantly and achieves an accuracy of 83.28%. The results demonstrated that the use of the genotype-guidance in the training

TABLE 3. The 74 genes selected by the intersection of the three methods (P-value < 0.02).

Gene list									
ABCC9	ANKLE2	ANKRD13D	ANKZF1	AP1G2	ARRDC2	ATAD2	ATG4B	ATIC	BARD1
BCL6	BMS1	BRAP	BTG2	CBX5	CD2AP	CEBPZ	CIRBP	CREBL2	CRK
CYB5A	CYBRD1	DCAF13	DCLRE1C	DDX42	DDX51	DENR	DHCR24	DNAJB6	DNPEP
DPYSL2	DROSHA	EMP2	ENY2	ERC1	ERCC6	ETNK1	FAM122B	FAM20B	FANCL
FOXK2	FRMD4B	GAK	GCN1L1	GGA3	GLUD1	GMDS	GTPBP2	GTPBP4	HIPK3
HIST1H1D	HIST1H4B	HIST1H4C	IL6ST	KIAA0368	LAMB2	LOC100133091	LPAR6	LRPPRC	MAPK14
MAT2A	MBOAT2	MCM4	MCM7	MYO10	NAA25	NOP56	PTGES3	RBMS3	RICTOR
SELENBP1	SNORA37	SPG7	TXNIP						

TABLE 4. The ablation study results of the proposed method for recurrence prediction.

	Handcrafted features	Deep features	Gene guidance	AUC	Specificity	Sensitivity	Accuracy
Model 1	x			0.6567	0.56	0.9	78.61%
Model 2		x		0.6714	0.59	0.89	79.09%
Model 3	x	x		0.7078	0.51	0.97	82.08%
Model 4 (GGR)	x	x	x	**0.7667**	**0.59**	**0.95**	**83.28%**

TABLE 5. The performance of GGR compared with the state-of-the-art methods.

Method		ACC	AUC	Specificity	Sensitivity
Wang, et.al. (2019) [12]	PCA+DT	61.36%	0.5716	0.45	0.69
Wang, et.al. (2019) [12]	PCA+SVM	67.05%	0.5833	0.31	0.85
Wang, et.al. (2019) [12]	PCA+RF	68.18%	0.6189	0.21	0.92
Lee, et. al. (2020) [14]	Relief-F + SVM	68.18%	0.5605	0.14	0.95
Lee, et. al. (2020) [14]	Relief-F + RF	67.04%	0.5745	0.28	0.86
Christie, et al. (2021) [17]	LASSO	61.36%	0.6756	0.62	0.71
Aonpong, et. al. (2020) [27]	F-test + ANN	78.61%	0.6567	0.56	0.9
Aonpong, et. al. (2020) [27]	ResNet50	79.09%	0.6714	0.59	0.89
Aonpong, et. al. (2020) [27]	DenseNet121	77.36%	0.6788	0.38	**0.97**
Marentakis, et al. (2021) [25]	Radiomics + deep learning	82.08%	0.7078	0.51	**0.97**
GGR (proposed method)	Gene guidance	**83.28%**	**0.7667**	**0.59**	0.95

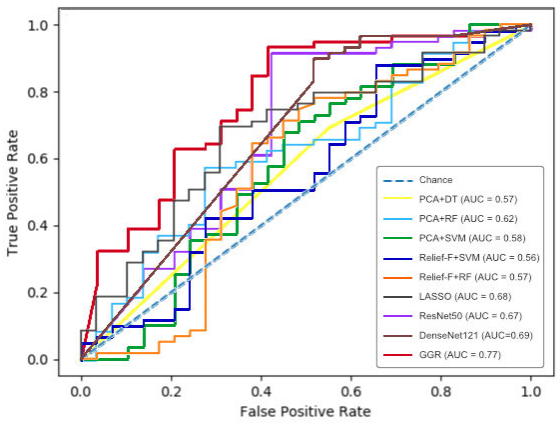

FIGURE 10. Average ROC of recurrence prediction using deep learning-based radiomics methods and the proposed GGR method.

phase is important to learn more useful features and enhance the CT-based recurrence prediction accuracy.

V. DISCUSSION AND CONCLUSION

In this paper, we studied the recurrence of NSCLC prediction using CT image analysis to help the doctor and patient to prepare for the risks that may occur. Since the traditional radiomics-based methods or recently proposed deep learning-based methods used only CT information, the prediction performance was limited. From the experiments, we can achieve 78.61% accuracy (AUC = 0.6567) for the best handcrafted based radiomics (F-test with ANN), 79.09% accuracy (AUC = 0.6714) for the deep learning-based (ResNet50) and 67.88% accuracy (AUC = 0.6788) for the deep learning-based (DenseNet121). By the proposed GGR, the prediction accuracy and AUC were significantly improved to 83.28% (AUC = 0.7667). The proposed prediction methods used genotype information to guide the model training, the prediction models can learn more useful information for early recurrence prediction and the proposed method only uses CT image for prediction in the test phase. We also performed experiments with real gene expression for early recurrence prediction. From our experiments, we can achieve 94.30% accuracy (AUC = 0.8915) for gene-expression analysis and 94.44% accuracy (AUC = 0.9158) for the fusing between gene-expression and CT radiomics signature. Both are much higher than the proposed method and can be considered as the upper limit performance of the proposed method. As our future work, we will improve the models and increase the training dataset to improve the prediction performance and make them closer to the genomics-based methods but use only the CT images.

REFERENCES

[1] M. A. Bareschino, C. Schettino, A. Rossi, P. Maione, P. C. Sacco, R. Zeppa, and C. Gridelli, "Treatment of advanced non small cell lung cancer," *J. Thoracic Disease*, vol. 3, no. 2, pp. 122–133, 2011.

[2] K. Zarogoulidis, P. Zarogoulidis, K. Darwiche, E. Boutsikou, N. Machairiotis, K. Tsakiridis, N. Katsikogiannis, I. Kougioumtzi, I. Karapantzos, H. Huang, and D. Spyratos, "Treatment of non-small cell lung cancer (NSCLC)," *J. Thoracic Disease*, vol. 5, no. 4, p. S389, 2013.

[3] V. Subramanian, M. N. Do, and T. Syeda-Mahmood, "Multimodal fusion of imaging and genomics for lung cancer recurrence prediction," in *Proc. IEEE 17th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2020, pp. 804–808.

[4] Y. Lu, R. Govindan, L. Wang, P.-Y. Liu, B. Goodgame, W. Wen, A. Sezhiyan, J. Pfeifer, Y.-F. Li, X. Hua, Y. Wang, P. Yang, and M. You, "MicroRNA profiling and prediction of recurrence/relapse-free survival in stage I lung cancer," *Carcinogenesis*, vol. 33, no. 5, pp. 1046–1054, May 2012.

[5] P. Thomas and L. Rubinstein, "Cancer recurrence after resection: T1 N0 non-small cell lung cancer," *Ann. Thoracic Surg.*, vol. 49, no. 2, pp. 242–247, Feb. 1990.

[6] P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R. G. P. M. van Stiphout, P. Granton, C. M. L. Zegers, R. Gillies, R. Boellard, A. Dekker, and H. J. W. L. Aerts, "Radiomics: Extracting more information from medical images using advanced feature analysis," *Eur. J. Cancer*, vol. 48, no. 4, pp. 441–446, Mar. 2012.

[7] V. Kumar, Y. Gu, S. Basu, A. Berglund, S. A. Eschrich, M. B. Schabath, K. Forster, H. J. W. L. Aerts, A. Dekker, D. Fenstermacher, D. B. Goldgof, L. O. Hall, P. Lambin, Y. Balagurunathan, R. A. Gatenby, and R. J. Gillies, "Radiomics: The process and the challenges," *Magn. Reson. Imag.*, vol. 30, no. 9, pp. 1234–1248, Nov. 2012.

[8] H. J. W. L. Aerts, E. R. Velazquez, R. T. H. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, F. Hoebers, M. M. Rietbergen, C. R. Leemans, A. Dekker, J. Quackenbush, R. J. Gillies, and P. Lambin, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature Commun.*, vol. 5, no. 1, pp. 1–9, Sep. 2014.

[9] R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: Images are more than pictures, they are data," *Radiology*, vol. 278, no. 2, pp. 563–577, Feb. 2016.

[10] S. S. Alahmari, D. Cherezov, D. B. Goldgof, L. O. Hall, R. J. Gillies, and M. B. Schabath, "Delta radiomics improves pulmonary nodule malignancy prediction in lung cancer screening," *IEEE Access*, vol. 6, pp. 77796–77806, 2018.

[11] S. Wold, E. Kim, and G. Paul, "Principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 2, nos. 1–3, pp. 37–52, 1987.

[12] X. Wang, H.-H. Duan, and S.-D. Nie, "Prognostic recurrence analysis method for non-small cell lung cancer based on CT imaging," *Proc. SPIE*, vol. 11321, Nov. 2019, Art. no. 113211T.

[13] I. Kononenko, "Estimating attributes: Analysis and extensions of RELIEF," in *Proc. Eur. Conf. Mach. Learn.* Berlin, Germany: Springer, 1994, pp. 171–182.

[14] S. Lee, J. Jung, H. Hong, and B. Kim, "Radiomic feature-based prediction model of lung cancer recurrence in NSCLC patients," *Proc. SPIE*, vol. 11515, Jun. 2020, Art. no. 115150N.

[15] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc., B, Methodol.*, vol. 58, no. 1, pp. 267–288, Jan. 1996.

[16] P. Aonpong, Q. Chen, Y. Iwamoto, L. Lin, H. Hu, Q. Zhang, and Y.-W. Chen, "Comparison of machine learning-based radiomics models for early recurrence prediction of hepatocellular carcinoma," *J. Image Graph.*, vol. 7, no. 4, pp. 117–125, 2019.

[17] J. R. Christie, M. Abdelrazek, P. Lang, and S. A. Mattonen, "A multimodality radiomics-based model for predicting recurrence in non-small cell lung cancer," *Proc. SPIE*. vol. 11600, Feb. 2021, Art. no. 116000L.

[18] T. A. D'Antonoli, A. Farchione, J. Lenkowicz, M. Chiappetta, G. Cicchetti, A. Martino, A. Ottavianelli, R. Manfredi, S. Margaritora, L. Bonomo, V. Valentini, and A. R. Larici, "CT radiomics signature of tumor and peritumoral lung parenchyma to predict nonsmall cell lung cancer postsurgical recurrence risk," *Academic Radiol.*, vol. 27, no. 4, pp. 497–507, Apr. 2020.

[19] Y. Zhou, L. He, Y. Huang, S. Chen, P. Wu, W. Ye, Z. Liu, and C. Liang, "CT-based radiomics signature: A potential biomarker for preoperative prediction of early recurrence in hepatocellular carcinoma," *Abdominal Radiol.*, vol. 42, no. 6, pp. 1695–1704, Jun. 2017.

[20] P. Starkov, T. A. Aguilera, D. I. Golden, D. B. Shultz, N. Trakul, P. G. Maxim, Q.-T. Le, B. W. Loo, M. Diehn, A. Depeursinge, and D. L. Rubin, "The use of texture-based radiomics CT analysis to predict outcomes in early-stage non-small cell lung cancer treated with stereotactic ablative radiotherapy," *Brit. J. Radiol.*, vol. 92, no. 1094, Feb. 2019, Art. no. 20180228.

[21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[23] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.

[24] D. Zhao, G. Xu, Z. Xu, T. Lukasiewicz, M. Xue, and Z. Fu, "Deep learning in computer-aided diagnosis and treatment of tumors: A survey," 2020, *arXiv:2011.00940*. [Online]. Available: http://arxiv.org/abs/2011.00940

[25] P. Marentakis *et al.*, "Lung cancer histology classification from CT images based on radiomics and deep learning models," *Med. Biol. Eng. Comput.*, vol. 59, no. 1, pp. 215–226, 2021.

[26] J. Lao, Y. Chen, Z.-C. Li, Q. Li, J. Zhang, J. Liu, and G. Zhai, "A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme," *Sci. Rep.*, vol. 7, no. 1, pp. 1–8, Dec. 2017.

[27] P. Aonpong, Y. Iwamoto, W. Wang, L. Lin, and Y. W. Chen, "Hand-crafted and deep learning-based radiomics models for recurrence prediction of non-small cells lung cancers," in *Innovation in Medicine and Healthcare*. Singapore: Springer, 2020, pp. 135–144.

[28] S. S. F. Yip and H. J. W. L. Aerts, "Applications and limitations of radiomics," *Phys. Med. Biol.*, vol. 61, no. 13, pp. R150–R166, Jul. 2016.

[29] E.-S. Lee *et al.*, "Prediction of recurrence-free survival in postoperative non–small cell lung cancer patients by using an integrated model of clinical information and gene expression," *Clin. Cancer Res.*, Vol. 14, no. 22, pp. 7397–7404, 2008.

[30] O. Gevaert, J. Xu, C. D. Hoang, A. N. Leung, Y. Xu, A. Quon, D. L. Rubin, S. Napel, and S. K. Plevritis, "Non–small cell lung cancer: Identifying prognostic imaging biomarkers by leveraging public gene expression microarray data—Methods and preliminary results," *Radiology*, vol. 264, no. 2, pp. 387–396, Aug. 2012.

[31] H. Liu, A. T. Kho, I. S. Kohane, and Y. Sun, "Predicting survival within the lung cancer histopathological hierarchy using a multi-scale genomic model of development," *PLoS Med.*, vol. 3, no. 7, p. e232, Jul. 2006.

[32] T. Shibata, S. Uryu, A. Kokubu, F. Hosoda, M. Ohki, T. Sakiyama, Y. Matsuno, R. Tsuchiya, Y. Kanai, T. Kondo, I. Imoto, J. Inazawa, and S. Hirohashi, "Genetic classification of lung adenocarcinoma based on array-based comparative genomic hybridization analysis: Its association with clinicopathologic features," *Clin. Cancer Res.*, vol. 11, no. 17, pp. 6177–6185, Sep. 2005.

[33] K. P. McCormick, R. W. Matthew, and C. M. Blake, "Experimental design, pre-processing, normalization and differential expression analysis of small RNA sequencing experiments," *Silence*, vol. 2, no. 1, pp. 1–19, 2011.

[34] S. Bakr *et al.*, "A radiogenomic dataset of non-small cell lung cancer," *Sci. Data*, vol. 5, no. 1, pp. 1–9, 2018.

[35] S. Bakr, O. Gevaert, S. Echegaray, K. Ayers, M. Zhou, M. Shafiq, H. Zheng, W. Zhang, A. Leung, M. Kadoch, J. Shrager, A. Quon, S. Plevritis, and S. Napel, "Data for NSCLC Radiogenomics Collection," *Cancer Imag. Arch.*, Oct. 2018, doi: 10.7937/K9/TCIA.2017.7hs46erv.

[36] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, and F. Prior, "The cancer imaging archive (TCIA): Maintaining and operating a public information repository," *J. Digit. Imag.*, vol. 26, no. 6, pp. 1045–1057, Dec. 2013.

[37] R. Sato, Y. Iwamoto, K. Cho, D.-Y. Kang, and Y.-W. Chen, "Accurate BAPL score classification of brain PET images based on convolutional neural networks with a joint discriminative loss function," *Appl. Sci.*, vol. 10, no. 3, p. 965, Feb. 2020.

[38] B. A. Simon, "Computed tomography studies of lung mechanics," *Proc. Amer. Thoracic Soc.*, vol. 2, no. 6, pp. 517–521, Dec. 2005.

[39] M. C. Van Verk, R. Hickman, C. M. J. Pieterse, and S. C. M. Van Wees, "RNA-seq: Revelation of the messengers," *Trends Plant Sci.*, vol. 18, no. 4, pp. 175–179, Apr. 2013.

[40] M. L. Gaddis, "Statistical methodology: IV. Analysis of variance, analysis of co variance, and multivariate analysis of variance," *Acad. Emergency Med.*, vol. 5, no. 3, pp. 258–265, Mar. 1998.

[41] H. Scheffe, *The Analysis of Variance*, vol. 72. Hoboken, NJ, USA: Wiley, 1999.

[42] S. E. Fienberg and H. O. Lancaster, "The chi-squared distribution," *Biometrics*, vol. 27, no. 1, p. 238, Mar. 1971.

[43] S. Rizzo, F. Botta, S. Raimondi, D. Origgi, C. Fanciullo, A. G. Morganti, and M. Bellomi, "Radiomics: The facts and the challenges of image analysis," *Eur. Radiol. Experim.*, vol. 2, no. 1, pp. 1–8, Dec. 2018.

[44] H. Kong, H. C. Akakin, and S. E. Sarma, "A generalized Laplacian of Gaussian filter for blob detection and its applications," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1719–1733, Dec. 2013.

[45] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.
[46] Y. Yu, H. Lin, J. Meng, X. Wei, H. Guo, and Z. Zhao, "Deep transfer learning for modality classification of medical images," *Information*, vol. 8, no. 3, p. 91, Jul. 2017.
[47] H. Y. Zhou, S. Yu, C. Bian, Y. Hu, K. Ma, and Y. Zheng, "Comparing to learn: Surpassing imagenet pre-training on radiographs by comparing image representations," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2020, pp. 398–407.
[48] D. M. Allen, "Mean square error of prediction as a criterion for selecting variables," *Technometrics*, vol. 13, no. 3, pp. 469–475, Aug. 1971.
[49] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proc. 19th Int. Conf. Comput. Statist.* Paris, France: Physica-Verlag HD, 2010, pp. 177–186.
[50] D. R. Cox, "The regression analysis of binary sequences," *J. Roy. Stat. Soc., B, Methodol.*, vol. 20, no. 2, pp. 215–232, Jul. 1958.
[51] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006.

PANYANAT AONPONG was born in Suphanburi, Thailand, in 1993. He received the B.Eng. degree in computer engineering from the King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand, in 2015, and the M.Eng. degree in information communication technology for embedded systems from Kasetsart University, Bangkok, Thailand, in 2017. He is currently pursuing the Ph.D. degree with the Department of Information Science and Engineering, Ritsumeikan University, Shiga, Japan. He is interested in computer image processing. He has experience in satellite imagery, when he was working with Geoinformatics Information and Space Technology Development Agency (GISTDA), Thai Government Organization, which works directly to the satellites and images.

YUTARO IWAMOTO (Member, IEEE) received the B.E., M.E., and D.E. degrees from Ritsumeikan University, Kusatsu, Japan, in 2011, 2013, and 2017, respectively. He is currently an Assistant Professor with Ritsumeikan University. His current research interests include image restoration, segmentation, classification of medical imaging, and deep learning

XIAN-HUA HAN (Member, IEEE) received the B.E. degree from Chongqing University, Chongqing, China, the M.E. degree from Shandong University, Jinan, China, and the D.E. degree from the University of Ryukyus, Okinawa, Japan, in 2005. From April 2007 to March 2013, she was a Postdoctoral Fellow and an Associate Professor with the College of Information Science and Engineering, Ritsumeikan University, Shiga, Japan. From April 2016 to March 2017, she was a Senior Researcher with the Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology, Japan. She is currently an Associate Professor with the Artificial Intelligence Research Center, Yamaguchi University, Japan. Her current research interests include image processing and analysis, pattern recognition, machine learning, computer vision, and hyperspectral image analysis. She is a member of the IEICE.

LANFEN LIN was born in Pingyang, China, in 1969. She received the B.S. and Ph.D. degrees in aircraft manufacture engineering from Northwestern Polytechnical University, in 1990, and 1995, respectively. She was a postdoctoral with the College of Computer Science and Technology, Zhejiang University (ZJU), China, from January 1996 to December 1997. She is currently a Full Professor with the College of Computer Science and Technology, ZJU, the Vice Director of the Artificial Intelligence Institute, ZJU, and a member of Advisory Expert Group for manufacturing industry informatization in Zhejiang. Her research interests include big data analysis, data mining, and knowledge management. She is currently leading a research group to exploring AI technologies for medical imaging. She is the leader of more than 40 projects, include the subproject of Major State Basic Research Program (973) of China, and the National High Technology Research and Development Program of China. She has published more than 140 research articles in the journals or international conference proceedings.

YEN-WEI CHEN (Member, IEEE) was born in Hangzhou, China, in 1962. He received the B.E. degree from Kobe University, Kobe, Japan, in 1985, and the M.E. and D.E. degrees from Osaka University, Osaka, Japan, in 1987 and 1990, respectively. He was a Research Fellow with the Institute of Laser Technology, Osaka, from 1991 to 1994. From October 1994 to March 2004, he was an Associate Professor and a Professor with the Department of Electrical and Electronic Engineering, University of the Ryukyus. He is currently a Professor with the College of Information Science and Engineering, Ritsumeikan University, Japan. He is also a Visiting Professor with the College of Computer Science and Technology, Zhejiang University, and Zhejiang Laboratory, Hangzhou. His research interests include medical image analysis and pattern recognition. He has published more than 300 research articles. He has received many distinguished awards. He is the Principal Investigator of several projects in bio-medical engineering and image analysis funded by Japanese Government.