Published in final edited form as:

Med Image Anal. 2018 December ; 50: 106–116. doi:10.1016/j.media.2018.09.004.

A pilot study using kernelled support tensor machine for distant failure prediction in lung SBRT

Shulong Li^a, Bin Li^a, Zhiguo Zhou^b, Ning Yang^c, Hongxia Hao^d, Michael R. Folkert^b, Puneeth Iyengar^b, Kenneth Westover^b, Hak Choy^b, Robert Timmerman^b, Steve Jiang^b, and Jing Wang^{b,*}

^aSchool of Biomedical Engineering, Guangdong Provincial Key Laboratory of Medical Image Processing, Southern Medical University, Guangzhou, 510515, China

^bDepartment of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, 75235, USA

^cDepartment of Medical Imaging, Guangdong No.2 Provincial People's Hospital, Guangzhou, 510317, China

^dSchool of Computer Science and Technology, Xidian University, Xi'an, 710071, China

Abstract

We developed a kernelled support tensor machine (KSTM)-based model with tumor tensors derived from pre-treatment PET and CT imaging as input to predict distant failure in early stage non-small cell lung cancer (NSCLC) treated with stereotactic body radiation therapy (SBRT). The patient cohort included 110 early stage NSCLC patients treated with SBRT, 25 of whom experienced failure at distant sites. Three-dimensional tumor tensors were constructed and used as input for the KSTM-based classifier. A KSTM iterative algorithm with a convergent proof was developed to train the weight vectors for every mode of the tensor for the classifier. In contrast to conventional radiomics approaches that rely on handcrafted imaging features, the KSTM-based classifier uses 3D imaging as input, taking full advantage of the imaging information. The KSTM-based classifier preserves the intrinsic 3D geometry structure of the medical images and the correlation in the original images and trains the classification hyper-plane in an adaptive feature tensor space. The KSTM-based predictive algorithm was compared with three conventional machine learning models and three radiomics approaches. For PET and CT, the KSTMbased predictive method achieved the highest prediction results among the seven methods investigated in this study based on 10-fold cross validation and independent testing.

Graphical Abstract

*Corresponding author. jing.wang@utsouthwestern.edu (Jing Wang).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

Support tensor machine; NSCLC; SBRT; Radiomics; Medical imaging

1. Introduction

Non-small cell lung cancer (NSCLC) is the most common type of lung cancer worldwide (World Health Organization, 2017). Approximately 15 – 20% of NSCLC patients present with early-stage, localized disease that can be controlled upon receiving treatment with curative-intent (Howlader N, 2016), although this number may rise with the recent implementation of CT-based screening programs (Team, 2013). In general, surgery has been the standard treatment for early stage NSCLC with long-term local control rates of 80–94% (Miller et al., 2016), but systemic failure rates range from 15–20% (Ginsberg et al., 1995; Martini et al., 1995). For patients who are not surgical candidates, stereotactic body radiation therapy (SBRT) has demonstrated local control rates over 95% (Nikolaev et al., 2016; Timmerman et al., 2010), becoming the standard of care for medically inoperable patients (Chi et al., 2010; Ettinger et al., 2010; Liu et al., 2015). Nevertheless, as with surgical patients, distant failure is common in this patient population. In the RTOG 0236 trial, the distant failure rate five years after SBRT was 31% (Timmerman et al., 2014). Another retrospective study on lung SBRT patients also found distant failure rates of 14.7% and 19.9% within two and five years, respectively (Senthil et al., 2012). Other prospective trials with small samples sizes also showed that SBRT for early-stage primary lung cancer achieved local control rates over 95% with high distant failure rates of about 29% (Nath et al., 2011). Unlike surgical cases, where features of the resected tumor specimens are the basis for identifying patients who may benefit from systemic therapy, risk models for identifying analogous patients in the SBRT population have not gained wide acceptance. Analogous to surgical patients, accurate prediction of early distant failure in NSCLC patients receiving SBRT has the potential to identify patients who would benefit from systemic therapy.

Medical imaging acquired as part of standard of care, such as computed tomography (CT) and fluorodeoxyglucose positron emission tomography (FDG-PET), has the potential to predict outcomes in early stage NSCLC treated with SBRT (Clarke et al., 2012; Lian et al., 2016; Lovinfosse et al., 2016; Tan et al., 2013). With recent advances in radiomics techniques, many approaches have been developed for predictive modeling based on medical images (Cook et al., 2014; Gillies et al., 2015; Kumar et al., 2012; Lambin et al., 2012;

Madabhushi and Lee, 2016; Tajbakhsh et al., 2016; Tan et al., 2013), mainly involving features/input determination and classifier learning.

In general, current methods of classification modeling based on medical images can be divided into three categories: 1) handcrafted feature-based approaches (Aerts et al., 2014; Lian et al., 2016; Liu et al., 2017; Namburete et al., 2015; Parmar et al., 2015; Vallières et al., 2015; Wimmer et al., 2016; Wu et al., 2016; Zhou et al., 2017); 2) feature learning-based approaches (Ciompi et al., 2015; Kooi et al., 2017; Kumar et al., 2015; Roth et al., 2016; Shen et al., 2015; Shin et al., 2016); and 3) voxel/pixel-based approaches (Huang et al., 2017; Khamis et al., 2017; Zuluaga et al., 2015).

In handcrafted feature-based methods, quantitative imaging features are manually extracted first. Feature selection methods are then used to select optimal feature subsets to reduce the redundancy of handcrafted features. Finally, the selected features are used to construct a predictive model. Although quantitative imaging-based features have shown to be effective in building predictive models, the generic nature of their features may not fully reflect the unique characteristics of a particular cancer (Chung et al., 2015). Moreover, different feature selection methods often obtain different optimal features subsets (Zhou et al., 2016), hindering the robustness of the classification.

Recently, feature learning-based approaches, such as the discovery radiomics method (Chung et al., 2015) and the deep convolutional neural network (CNN) (Shin et al., 2016) have been developed using the intrinsic features learned automatically for the predictive model. These methods require large scale annotated datasets (such as ImageNet (Krizhevsky et al., 2012)) to achieve acceptable prediction accuracy. However, obtaining such a large annotated dataset for radiation treatment outcome prediction is still challenging (Miotto et al., 2017).

Investigators have proposed constructing models using original images without handcrafted features and feature learning, i.e., voxel/pixel-based models. The conventional voxel/pixel-based approaches first vectorize the image and then use the vectorization data as input for modeling. In this case, the dimensionality of the input sample is usually high. High space dimensionality may induce small sample size problems (Raudys and Jain, 1991) and the curse of dimensionality (Keogh and Mueen, 2011) for medical data with small training sample size, leading to overfitting and decreasing classification accuracy. Dimensionality reduction methods, such as the down-sample method (Khamis et al., 2017; Magnin et al., 2009), were developed to improve the performance of classifiers. However, image vectorization with an intrinsic tensor structure breaks the geometry structure and correlation in the original image so that its information cannot be fully taken into consideration. Meanwhile, reducing dimensionality also induces the loss of available tumor information.

In contrast to the methods mentioned earlier, the tensor space model (Kolda and Bader, 2009; Liu et al., 2010; Shashua and Levin, 2001) uses high order imaging with an intrinsic high order tensor structure as input for modeling, eliminating the need for handcrafted imaging features, feature learning, or image vectorization. Three dimensional (3D) medical imaging methods, such as PET and CT, have an intrinsic tensor structure suitable for tumors

as 3D objects. Support tensor machine (STM) is a common tensor space model that is the generalization of support vector machine (SVM) from vector space to tensor space. In contrast to conventional radiomics approaches that rely on handcrafted imaging features, the STM-based classifier uses 3D imaging as input for the model, taking full advantage of the information from whole images without using handcrafted features. The STM-based classifier does not break the natural 3D geometry structure of the image and correlation in the original image.

Tao et al. introduced first the STM framework based on the multilinear algebra theory with a brief discussion on potential kernelization (Tao et al., 2005). They further gave the alternating projection optimization procedure with a convergent proof for STM without kernelization (Tao et al., 2007). This algorithm and its developed forms have been applied successfully to pedestrian detection, face recognition, and remote sensor (Biswas and Milanfar, 2016; Hao et al., 2013; Zhang et al., 2011). In most of these applications, STM models are trained by either iterative projection techniques (Zhang et al., 2011) or rank decomposition methods (Hao et al., 2013) because the training model of STM is not a convex optimization problem in tensor space and its closed form solution is unavailable. In these training algorithms, the classification hyper-planes are trained in the original tensor space. Thus these STMs are called as linear STMs (LSTM) since they only involve multilinear operations without the transformation from original tensor space to the feature tensor space. However, similar to kernel-based SVM models, it could be beneficial to construct the STM model in an adaptive feature tensor space transformed from original tensor space. Motivated by the application of kernels in SVM, we develop a kernelled STM algorithm (KSTM) for the STM in an adaptive feature tensor space that uses 3D tumor tensors as input to predict distant failure in early stage NSCLC treated with definitive SBRT in this pilot study. The KSTM iterative algorithm with a convergence proof is given and used to train the weight vectors for each tensor mode for the classifier.

2. Materials and Methods

2.1. Patient cohort

Our study included a cohort of 110 patients with inoperable early stage (IA and IB) NSCLC. They had been treated with SBRT from 2006 to 2016 with a median follow-up of 18 months. Twenty-five (22.73%) of these patients experienced distant failure. For each patient, pre-treatment CT and PET images were acquired by one of three clinical scanners: the SIEMENS Biograph 64/1094 (Siemens Medical Solution, Malvern, PA), the Philips Gemini TF/Dual GS (Philips Healthcare, Andover, MA) or the GE Discover ST (GE Healthcare, Waukesha, WI). The CT volume was composed of 274 to 355 slices with 512×512 pixels. The PET volume was composed of 274 to 355 slices with 168×168 pixels, 144×144 pixels or 128×128 pixels.

2.2. Tensor Construction

Tumor tensors were constructed from both PET and CT images. The standard uptake value (SUV) (Vansteenkiste et al., 1999) was first calculated for each PET image. Tumors were then segmented from both PET and CT images in a semi-automatic way as follows. The

middle slice was segmented using the object information based interactive segmentation method (OIIS) that requires a seed region of tumor identified manually (Zhou et al., 2013). Once the central slice was segmented, the other slices were segmented by the well-known OTSU method (Otsu, 1979) that considers the similarity between two adjacent slices. For each imaging modality after segmentation, the 3D tumor tensors were constructed with the same size that was the biggest value among all tumors sizes, which center the tumors were located in and which periphery zeros were filled in. The tensor size of PET imaging was $17 \times 17 \times 17$ and the tensor size of CT imaging was $53 \times 56 \times 20$.

All tumor ROIs were interpolated to the same resolution using a 3D spline interpolation method (Sathyamurthy and Raff, 1975). To determine the tensor size after interpolation, the product between tumor ROI size (number of voxels) and physical size of each voxel at the same dimension was calculated. The quotient of the product was then divided by the physical size of voxel interpolated tumor ROI at its corresponding dimension became the size of the interpolated tumor. After interpolation, the tensor size of PET imaging was $34 \times 34 \times 43$ and the tensor size of CT imaging was $55 \times 56 \times 80$.

Consider the noise from interpolation processing, the three-dimensional Gabor filter banks were used to denoise the images (Ahmed, 2011) by obtaining the frequency information of the local space domain, which are similar to that of the human visual system. The four constructed tumor tensors for PET imaging and CT imaging without interpolation and Gabor filtering are illustrated in Fig. 1. The tensors with 3D spline interpolation and 3D Gabor filtering were used as the input of the KSTM algorithm.

2.3. Kernelled Support Tensor Machine (KSTM)

The transformation from original tensor space $R^{J_1 \times J_2 \times \dots \times J_N}$ to an adaptive feature tensor space $R^{I_1 \times I_2 \times \dots \times I_N}$ is defined as follows

$$\Phi: R^{J_1 \times J_2 \times \dots \times J_N} \rightarrow R^{I_1 \times I_2 \times \dots \times I_N}$$

satisfying

$$\Phi(X) = \left(\Phi_{i_1, i_2, \dots, i_N}^{(X)} \right)_{I_1 \times I_2 \times \dots \times I_N},$$

where $X \in R^{J_1 \times J_2 \times \dots \times J_N}$.

For tensor $X \in R^{J_1 \times J_2 \times \dots \times J_N}$, a hyper-tensor plane is constructed in the adaptive feature tensor space $R^{I_1 \times I_2 \times \dots \times I_N}$ based on the mode product (De Lathauwer, 1997) between tensor and vector:

$$\mathcal{O}(X) \prod_{n=1}^N \times_n w^{(n)} + b = 0, \quad (1)$$

where $w^{(n)} \in \mathbb{R}^{I_n}$, $n = 1, 2, \dots, N$ is a weight vector for mode n and b is the bias to be determined. A visual illustration of the mode product between a 3D tensor and three vectors is presented in Fig. 2.

Given the training samples set $\left\{X_m \in \mathbb{R}^{J_1 \times J_2 \times \dots \times J_N}, y_m \in \left\{-1, 1\right\}\right\}_{m=1}^M$, where y_m is the class label of sample X_m , the hyper-tensor plane (1) is determined by solving the following optimization problem the similar to linear STM:

$$\operatorname{argmin} J(w^{(n)}(n = 1, 2, \dots, N), b, \xi_m(m = 1, 2, \dots, M)) = \frac{1}{2} \prod_{n=1}^N \|w^{(n)}\|_F^2 + C \sum_{m=1}^M \xi_m, \quad (2)$$

subjected to

$$y_m(\mathcal{O}(X_m) \prod_{n=1}^N \times_n w^{(n)} + b) \geq 1 - \xi_m, \quad (3)$$

$$\xi_m \geq 0, m = 1, 2, \dots, M, \quad (4)$$

where ξ_m is the misclassification error of the m th training sample X_m ($m = 1, 2, \dots, M$), and C is the trade-off between the classification margin and the misclassification error.

Because the optimization problem defined above is not convex in the linear space, a closed form solution is unavailable. Moreover, it is hard to give the particular transformation $\mathcal{O}$ since it is adaptive. Motivated by the iterative procedure of LSTM and SVM solver with kernel, a kernelled STM iterative algorithm (KSTM) is proposed by orderly fixing $N - 1$ weight vectors in (2)–(4)–(3) inherited from the previous step. This alternating approach is equivalent to the following N training models of SVM in feature space ($n = 1, 2, \dots, N$):

$$\operatorname{argmin} J(w^{(n)}, b^{(n)}, \xi_m^{(n)}(m = 1, 2, \dots, M)) = \frac{1}{2} \|w^{(n)}\|_F^2 \prod_{l=1}^{l \neq n} \|w^{(l)}\|_F^2 + C \sum_{m=1}^M \xi_m^{(n)} \quad (5)$$

subjected to

$$y_m \left((w^{(n)})^T \left(\mathcal{O}_n \left(X_m \prod_{l=1}^{l \neq n} \times_l P_l(w^{(l)}) \right) \right) + b^{(n)} \right) \geq 1 - \xi_m^{(n)}, \quad (6)$$

$$\xi_m^{(n)} \geq 0, m = 1, 2, \dots, M, \quad (7)$$

where $w^{(l)}$, $l = 1, 2, \dots, N$ and $l \neq n$ are fixed, $\varnothing_n: R^n \rightarrow R^n$ is an adaptive projective mapping from R^n to R^n and $P_l: R^l \rightarrow R^l$ is a projective mapping from R^l to R^l defined as (8) and (12).

The detailed iterative procedures to solve equations (5)–(7) are described in Table 1.

Once the iterative KSTM training model defined in equations (5)–(7) is solved, the classification hyper-tensor plane is fixed and the class label of tensor sample X can be predicted using the following classification function:

$$f(X) = \text{sign}(\sum_{i=1}^M \alpha_i^{(n)} y_i K(Y_i^{(n)}, X \prod_{1 \leq l \leq N, l \neq n} \times_l w^{(l)}) + b^{(n)}), \quad (14)$$

where the kernel function K is optional, which is similar to the kernel function applied in SVM. The radial basis function (RBF) kernel has demonstrated superior performances in various applications. Thus our proposed algorithm also coupled with the RBF kernel in our all experiments.

The convergence of the iterative procedure is proved below.

Theorem 1. The iterative procedure (5)–(7) will monotonically decrease the objective function values in the optimization problem defined in equations (2)–(4), converging the KSTM iterative algorithm (5)–(7).

Proof: Equation (2) can be rewritten as:

$$J(w^{(1)}, w^{(2)}, \dots, w^{(N)}) = \frac{1}{2} \prod_{n=1}^N \|w^{(n)}\|_F^2 + C \sum_{m=1}^M \xi_m. \quad (15)$$

First, we set the initial weight vectors $w^{(l)}$ and projective mapping P_l in equation (5)–(7) in Step 1 and fix $w^{(l)}$ in Step 2 of Table 1. Without loss of generality, starting $n = 1$, we assume $Y_i^{(n)} = X_i \prod_{1 \leq l \leq N, l \neq n} \times_l P_l(w^{(l)})$, $i = 1, 2, \dots, M$. At the point of the minimum on the Lagrangian of the optimization problem (5)–(7), the one order derivative of this Lagrangian with respect to $w^{(1)}$ must be zero. Thus

$$w^{(1)} = \sum_{i=1}^M \alpha_i^{(1)} y_i \varnothing_1(Y_i^{(1)}), n = 2, 3, \dots, N. \quad (16)$$

and the optimization problem (5)–(7) is equivalent to its dual problem defined in (9)–(11) with a kernel function K (Cortes and Vapnik, 1995). We then solve the quadratic program

problem (9)–(11) to obtain the unique solution $\alpha_i^{(1)}$ ($i = 1, 2, \dots, M$) and update $w^{(1)}$ with $w_1^{(1)} = \sum_{i=1}^M \alpha_i^{(1)} y_i \phi_1(Y_i^{(1)})$. Second, set $P_1(w^{(1)}) = \sum_{i=1}^M \alpha_i y_i Y_i^{(1)}$ and $Y_i^{(2)} = X_i \prod_{1 \leq l \leq N} \times_l P_l(w^{(l)})$, $i = 1, 2, \dots, M$ with the newest $P_1(w^{(1)})$. Input $P(w^{(l)})$ 1 I $N, I = 2$ to the optimization problem (5)–(7) to obtain $\alpha_i^{(2)}$ ($i = 1, 2, \dots, M$) by solving its dual problem (9)–(11). Update $w^{(2)}$ with $w_1^{(2)} = \sum_{i=1}^M \alpha_i y_i \phi_2(Y_i^{(2)})$. We repeat the second steps above until we obtain the last unique solution $\alpha_i^{(N)}$ ($i = 1, 2, \dots, M$) and update $w^{(N)}$ with $w_1^{(N)} = \sum_{i=1}^M \alpha_i y_i \phi_N(Y_i^{(N)})$ in the first iteration. We can then obtain the weight vectors $\{w_1^{(n)} | n = 1, 2, \dots, N\}$ in the first iteration. For the second iteration, we repeat these steps in the first iteration by using $\{P_n(w_1^{(n)}) | n = 1, 2, \dots, N\}$ as fixed weight vectors, and we obtain weight vectors $\{w_2^{(n)} | n = 1, 2, \dots, N\}$ in the second iteration. As we continue the iterations process, a sequence of weight vectors $\{w_t^{(n)} | n = 1, 2, \dots, N\}_{t=1}^{\infty}$ is obtained, where t is the iterative number. Since (9)–(11) is a quadratic program problem $\alpha_i^{(n)}$ ($i = 1, 2, \dots, M$) is the globally optimal solution to the corresponding maximum dual optimization problem (9)–(11) which is equivalent to the optimization problem (5)–(7). Thus $w_i^{(n)}$ is the globally optimal solution to the corresponding minimum optimization problem. Thus,

$$\begin{aligned} J(w_1^{(1)}, w_0^{(2)}, w_0^{(3)}, \dots, w_0^{(N)}) &\geq J(w_1^{(1)}, w_1^{(2)}, w_0^{(3)}, \dots, w_0^{(N)}) \geq \dots \geq J(w_1^{(1)}, w_1^{(2)}, w_1^{(3)}, \dots, w_1^{(N)}) \geq \\ J(w_2^{(1)}, w_1^{(2)}, w_1^{(3)}, \dots, w_1^{(N)}) &\geq J(w_2^{(1)}, w_2^{(2)}, w_1^{(3)}, \dots, w_1^{(N)}) \geq \dots \geq J(w_2^{(1)}, w_2^{(2)}, w_2^{(3)}, \dots, w_2^{(N)}) \geq \\ &\dots \geq \\ J(w_t^{(1)}, w_{t-1}^{(2)}, w_{t-1}^{(3)}, \dots, w_{t-1}^{(N)}) &\geq J(w_t^{(1)}, w_t^{(2)}, w_{t-1}^{(3)}, \dots, w_{t-1}^{(N)}) \geq \dots \geq J(w_t^{(1)}, w_t^{(2)}, w_t^{(3)}, \dots, w_t^{(N)}) \geq \\ &\dots \geq 0. \end{aligned}$$

Therefore, $\{J(w_t^{(1)}, w_t^{(2)}, w_t^{(3)}, \dots, w_t^{(N)})\}_{t=1}^{\infty}$ is a monotonically decreasing sequence with down boundary, and therefore converges. #

The change in function values of the left formula in equation (13) was plotted with different iteration numbers for PET and CT, in Fig. 3 (a) and (b), respectively. The function value decreased rapidly to zero within a few iterations, illustrating the fast convergence of the STM iterative algorithm (Fig. 3).

2.4. Three dimensional data balancing

Class imbalance had to be addressed since distant failure occurred in a minority of cases. The Synthetic Minority Over-sample Technique (SMOTE) (Chawla et al., 2002) is a common method to solve the class imbalance problem. It can augment the decision region of the minority class by generating a synthetic sample based on minority class information. Typically, SMOTE is produced in the vector space using the K-nearest neighborhood (KNN) graph based on Euclidean distance. However, because we used 3D imaging as input for the STM-based algorithm, the conventional SOMTE is no longer applicable. Thus, we propose a

high order SMOTE by using the KNN graph in the tensor space based on the tensor distance defined as follows:

$$d(X, Y) = \|X - Y\|_F = \sqrt{\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} (x_{i_1, i_2, \dots, i_N} - y_{i_1, i_2, \dots, i_N})^2}. \quad (17)$$

This high order SMOTE can generate synthetic tensor samples, but not vector samples, based on minority class information.

3. Experimental Setup

3.1. Workflow of the KSTM-based classifier

The overall workflow of the KSTM-based classifier is illustrated in Fig. 4. SUVs are first calculated for PET. For each patient, the primary tumors are segmented semi-automatically on both pre-treatment CT and PET. Three-dimensional tumor tensors for each imaging modality are constructed and used as input for the KSTM-based classifier. The 3D spline interpolation is used to equalize the resolution and the 3D Gabor filter is applied to filter the noise. High order SMOTE is used to balance the training samples with tensor structure. A KSTM iterative algorithm is used to train the weight vectors for every mode of the tensor for the classifier.

3.2. Multi-modality fusion for PET and CT

As PET and CT measure different intrinsic characteristics of a tumor, the combination of information learned from PET and CT could improve the predictive performance. In this work, we combined the output scores (i.e. the binary probabilities) of the KSTM-based classifier for PET (KSTM-PET) and the KSTM-based classifier CT (KSTM-CT) using a reliable classifier fusion (RCF) strategy (RCF-PET-CT) (Zhiguo Zhou, 2017). The RCF strategy does not only consider the relative importance of output scores from different modalities, but also takes the reliability of output scores from different modalities to get more reliable fusion results.

3.3. Methods for comparative testing

The comparative methods include three conventional machine learning methods (SVM, SVM-PCA and LSTM) and three conventional radiomics methods (SVM-ReliefF, SVM-SFS and RF-WLCX).

1) SVM: As the basic model of STM, SVM with vectorization of the tumor image as input for modeling was compared with our proposed algorithm. Performance was improved using the 3D spline interpolation to equalize the resolution and the 3D Gabor filtering to denoise on both PET and CT imaging. We used vectorization of the tumor image after 3D spline interpolation and 3D Gabor filtering as input for the SVM-based classifier with radial basis function (RBF) kernel.

2) SVM-PCA: Considering the high dimensionality of the image vectorization, principal component analysis (PCA) (Abdi and Williams, 2010) was used to reduce sample dimensionality and extract the principal components. The number of the principal components was selected using the smallest size with a variance contribution rate of 100% (PET: 44; CT: 47). The principal components extracted by PCA from vectorization of the tumor image with 3D spline interpolation and 3D Gabor filtering were used as input for the SVM-based classifier with radial basis function (RBF) kernel.

3) LSTM: Linear support tensor machine constructs the classification model in original data tensor space. It is a simplified form of KSTM, thus it is compared with the proposed KSTM algorithm. The tensors after 3D spline interpolation and 3D Gabor filtering are used as the input for modeling.

4) SVM-ReliefF: Conventional radiomics methods referring to handcraft-based extraction and analysis of high-throughput quantitative features mainly include three steps: 1) feature extraction; 2) feature selection; 3) classification. In this method, 240 texture features were extracted based on the Grey Level Concurrence Matrix (GLCM) (Davis et al., 1979). In particular, for each distance from {1, 2, 4, 8} and each gray level from {8, 16, 32, 64, 128}, twelve texture features (Zhou et al., 2017) were calculated based on the averaging GLCM at thirteen directions. Nine intensity features based on the intensity histogram, and eight geometry features associated with tumors were also extracted. Furthermore, 257 wavelet features were calculated using a coiflet transformation (Parmar et al., 2015), which were the transformed domain representation of the intensity and textural features. A total of 514 radiomic features were used in this analysis. The classical feature selection method ReliefF (Kononenko et al., 1997) was used to select the optimal feature subset based on the training dataset. The selected feature subset on the training dataset was used as the input for the SVM model with a RBF kernel to train the classification model by the 10-fold cross validation method. The selected feature subset on the testing dataset was used to independently evaluate the performance of the trained classification model.

5) SVM-SFS: As the traditional feature selection method, the sequential forward feature selection (SFS) method (Kohavi and John, 1997) coupled with the SVM classifier using the above 514 handcrafted features as input was also compared with our proposed algorithm. SFS with the minimal classification error as criterion based on the quadratic discrimination was used to select the optimal feature subsets with 10-fold cross validation method in the training dataset. We obtained 20 optimal feature subsets after the 20 times of feature selection process. The frequency of each feature appearing in these 20 subsets was then calculated and sorted decreasingly. Let $\{p_i | i = 1, 2, \dots, n\}$ denote the set consisting of such frequencies. Subsequently, new feature subsets are constructed by collecting the features with a frequency higher than the value $p_i (i = 1, 2, \dots, n)$ in the training dataset. Then, n new feature subsets and n groups of the predictive results can be obtained by using the new feature subset as input for the SVM-based classifier with the 10-fold cross validation method. The best results are selected from the n groups of predictive results, and the corresponding optimal features subset is obtained, which was also used as the optimal features subset in the independent testing.

6) RF-WLCX: The 514 features extracted as above (SVM-ReliefF) was used here. Parmar et al. examined twelve classification methods and fourteen feature selection methods (Parmar et al., 2015). They found that the Wilcoxon test based feature method (WLCX) and the classification method random forest (RF) had the highest prognostic performance. Thus we used WLCX to rank the importance for each feature. Subsequently, we selected the feature subset consisting of the top important features with different number from {5, 10, ..., 100} and obtain 20 feature subsets. The RF classifier was used to perform the classification with the selected feature subset as input in the training dataset by the 10-fold cross validation method. Then 20 results can be obtained for 20 selected feature subsets. The best results were selected from the 20 results and the corresponding optimal features subset is obtained, which was also used as the optimal features subset in independent testing.

3.4. Experimental settings

Firstly, 33 patients (30% over all participants) were selected randomly as an independent testing dataset, in which 7 participants experienced distant failures. The remaining patients were used for model training and validation, in which 18 patients experienced distant failure. A 10-fold cross validation method was employed to evaluate performance in the training dataset. All patients in the training dataset were randomly partitioned into 10 subsets with a size either 7 or 8. For each experimental round in the training processing, one subset was designated as the testing data and the remaining subsets as the training data. This process was repeated 5 times to avoid any bias introduced by random data partitioning. The optimal parameter group was selected and the corresponding classification results were obtained by averaging the results 5 times. An independent testing method was used to evaluate performance in the testing dataset with the model trained by the training dataset.

For the KSTM and SVM algorithms, the trade-off parameter C was chosen from $\{2^{-20}, 2^{-19}, \dots, 2^{20}\}$ and the free parameter in the RBF kernel was also chosen from $\{2^{-20}, 2^{-19}, \dots, 2^{20}\}$. For the LSTM algorithm, the trade-off parameter was also chosen from $\{2^{-20}, 2^{-19}, \dots, 2^{20}\}$. For the random forest algorithm, the number of tree was chosen from $\{250, 255, 260, \dots, 600\}$. The threshold in the stopping condition (13) was empirically set to 10^{-2} .

We adopted four metrics as evaluation criteria, including the area under the receive operating characteristic curve (AUC) (Fletcher et al., 2012), classification accuracy (ACC), sensitivity (SEN), and specificity (SPE). Let TP, TN, FP, and FN indicate true positive rate, true negative rate, false positive rate, and false negative rate, respectively. These evaluation metrics are defined as $ACC = (TP + TN)/(TP + TN + FP + FN)$, $SEN = TP/(TP + FN)$, $SPE = TN/(TN + FP)$.

4. Results

4.1. Performance with 3D interpolation and 3D Gabor filtering

We use the results of 10-fold cross validation in training dataset to evaluate the performance with 3D interpolation and 3D Gabor filtering. The performance is presented in Fig. 5. The KSTM-based classifier coupled with 3D interpolation and 3D Gabor filtering (KSTM-IN3-Gabor) consistently outperformed the KSTM-based classifier without interpolation and

Gabor filtering, and the KSTM-based classifier coupled with 3D interpolation (KSTM-IN3) for both PET and CT. KSTM-IN3 consistently outperformed KSTM except for sensitivity in PET. The corresponding ROC curves for the KSTM-based classifier are shown with and without 3D spline interpolation and 3D Gabor filtering in a one-time 10-fold cross validation experiment (Fig. 6). Also, KSTM-IN3-Gabor outperformed KSTM-IN3, and KSTM-IN3 outperformed KSTM. Thus, the tensors with 3D spline interpolation and 3D Gabor filtering were used as the input for the KSTM algorithm.

4.2. Comparison with conventional methods

The results of the comparative testing of seven classification algorithms are shown in table 2 for PET imaging and table 3 for CT imaging. For PET in 10-fold cross validation on training dataset, the KSTM-based predictive algorithm achieved the highest AUC (0.81), accuracy (76.10%), sensitivity (78.89%) scores among the seven methods investigated in this study, although its specificity score (75.25%) was slightly lower. For PET on the independent testing dataset, the KSTM-based predictive algorithm also achieved the highest AUC (0.74), accuracy (75.76%) and sensitivity (71.43%) scores, although specificity score (76.92%) was lower than LSTM that had imbalanced sensitivity (42.86%) and specificity (80.77%). For CT, the KSTM-based predictive algorithm also achieved the highest AUC (0.81), accuracy (78.44%), sensitivity (77.78%) scores in 10-fold cross validation on training dataset as well as the highest AUC (0.80), accuracy (75.76%), sensitivity (71.43%) and specificity (76.92%) scores in independent dataset among the seven methods investigated in this study.

The corresponding ROC curves for the seven predictive algorithms are shown in a one-time 10-fold cross validation experiment in training dataset and a one-time independent testing experiment in testing dataset (PET: Fig. 7; CT: Fig. 8), indicating that the proposed KSTM-based algorithm coupled with 3D spline interpolation and 3D Gabor filtering outperformed the other six methods.

We used the unpaired t-test with a 95% confidence interval to assess whether the difference in performance between the KSTM-based algorithm and each competing method was statistically significant in 10-fold cross validation. The corresponding p-values are reported in Table 4 for training results. These results show that our KSTM-based algorithm is significantly better than the other methods, as shown by the p-values < 0.05.

4.3. Multi-modality fusion of PET and CT based on RCF

The results of the fusion of PET and CT based on RCF strategy (RCF-PET-CT) on the independent testing dataset were summarized in Table 5. Compared with the results from KSTM-based classifier using PET or CT imaging only as the input, RCF-PET-CT obtained higher AUC (0.84), accuracy (84.85%) and specificity (88.46%) and the same sensitivity (71.43%). The ROC curves of the KSTM- PET and KSTM-CT, and RCF-PET-CT shown in Fig. 9 also demonstrate that the RCF-PET-CT had a better performance.

5. Discussion

We evaluated an KSTM-based classifier that uses 3D tumor images as input to predict distant failure in early stage NSCLC patients treated with SBRT. The experimental results

showed that the KSTM-based predictive algorithm outperforms traditional machine learning approaches and conventional radiomics approaches. It achieved the following better metrics on CT than PET for AUC, accuracy, sensitivity, and specificity: 0.81, 78.44%, 77.78%, and 78.64% in 10-fold cross validation, respectively. It also achieved the following better metrics on CT than PET for AUC, accuracy, sensitivity, and specificity: 0.80, 75.76%, 71.43%, 76.92% for CT in the independent testing, respectively. These results have implications for patients with early stage NSCLC who might benefit from the addition of systemic therapy based on a calculated increased risk for distant failure. These short-term predictive models will be useful for stratifying patients for risk-adapted prospective therapeutic protocols.

The proposed model can fully characterize the tumor properties through tensor structure, potentially overcoming the limitations of conventional methods that rely on handcrafted features. The KSTM-based classifier takes full advantage of the image information from whole images without using handcrafted features. Additionally, for conventional radiomics approaches, the feature selection is needed to reduce the redundancy of handcrafted features. However, the same feature selection method coupled with different classifiers or the same classifier coupled with different feature selection methods often yield different selection results, affecting the robustness of the classification results. Therefore, the KSTM-based algorithm is more robust than conventional radiomics approaches.

Compared with the voxel/pixel-based approaches that are restricted to one-dimensional vectors unfolded from tensors as input for modeling, the KSTM-based classifier learns fewer parameters compared to the SVM-based models. The KSTM model involves a reduced number of parameters for optimization. SVM with vectorization of the image as input needs learning $\left(\prod_{n=1}^N I_n + M + 1\right)$ parameters for the training samples mentioned in subsection 2.3.

However, the STM-based classifier with the high order image tensor as input only needs learning $\left(\sum_{n=1}^N I_n + M + 1\right)$ parameters, which are less than $\left(\prod_{n=1}^N I_n + M + 1\right)$, the number of parameters learned in the SVM model. Additionally, the weight vector dimensionality in the KSTM iterative algorithm (13)–(15) is I_n ($n = 1, 2, \dots, N$), which is lower than the dimensionality $\prod_{n=1}^N I_n$ of the SVM weight vector with vectorization of the image as input. For example, if the original tensor space is projected to the feature tensor space with the same dimensionality, the number of parameters is calculated as $55 + 56 + 80 + 48 + 1 = 239$ and the dimensionality of the weight vectors in the STM iterative process is 55, 56, 80 in the KSTM-IN3-Gabor algorithm for CT imaging (the tensor size is $55 \times 56 \times 80$, as seen in section 2.2). The number of parameters is $55 * 56 * 80 + 48 + 1 = 246449$ and the dimensionality of the weight vector is $55 * 56 * 80 + 246400$ in the algorithm SVM-IN3-Gabor for CT imaging.

The KSTM-based classifier holds all imaging information without dimensionality reduction usually needed for the voxel/pixel based approaches. Meanwhile, the KSTM-based classifier maintains the natural 3D geometry structure of the image and correlation in the original image. Additionally, the KSTM based algorithm trains the classification model on the adaptive feature tensor space based on the kernel function. It improves the performance compared to the LSTM based algorithm that trains the classification model on original tensor space. Although only 3D KSTM based algorithm is used in this study, the proposed

algorithm can solve the classification problem with any dimensionality. The KSTM based algorithm is applicable to the dataset with small scale. This is in contrast to feature learning method such as deep learning, which in general requires a large scale dataset to learn features effectively for classification.

In this study, the tensors as the input of the proposed algorithm were constructed based on the tumor ROI segmented using a semi-automatic way. As the KSTM does not require the calculation of texture or geometric features explicitly, we expect the segmentation accuracy of tumor will not dramatically affect the results as compared to the standard radiomic approaches. On the other hand, even though KSTM uses the original image (after segmentation) as the input, it does not involve auto-feature extraction process that is typically involved in convolutional neural network. Therefore, a large volume that involves too much background (e.g., whole lung slice) will adversely affect the performance of KSTM. Nevertheless, a region of interest (ROI) covering the tumor with minimal involvement of background can be potentially used as the input in KSTM. The influence of segmentation as well as input ROI is worthy of investigation in a future study.

The 3D spline interpolation was used to obtain the tensor with the same resolution along different directions and the three-dimensional Gabor filter banks were used to denoise the images in this work. Experimental results in Figures 5 and 6 demonstrate that both 3D interpolation and noise filtering can improve the performance of KSTM. However, different interpolations schemes and noise reduction filters may potentially affect the performance of KSTM, which requires further investigation.

It is noted that pre-treatment PET and CT images were acquired at different hospitals even though patients received SBRT treatments from a single institute. Thus acquisition protocol variations exist among images we used in this study. As shown in many radiomic approaches, some radiomic features can be sensitive to scanning protocols. The KSTM method used in this work does not require the calculation of features explicitly, thus it may be less sensitive to imaging protocol variation as compared to other radiomics approaches that rely on extracted quantitative features. Nevertheless, the performance of KSTM may be still affected by scanning protocol or acquisition parameter variations as reconstructed PET/CT images are used as the input for KSTM. However, due to a relatively small patient size we currently have, we are not able to quantify the influence of PET and CT scanning protocols to the performance of KSTM by further grouping patients according to scanning protocols. If the scanning protocols and acquisition parameters are harmonized for PET and CT, the performance of KSTM could be further improved.

The proposed KSTM-based classifier uses the KSTM iterative procedure (Table 1) to train the weight vectors for every mode of the tensor for the classifier. We proved the convergence of the KSTM iterative procedure, although the solution may not necessarily be optimal globally. Because the optimization problem defined by equations (2)–(4) is not convex in the linear space, a closed form solution by the linear algebra theory is unavailable. The convex optimization theory (Li et al., 2009) on Riemannian manifolds may be used to study the closed form solution by constructing a Riemannian manifold and proving that the optimization problem defined by equations (2)–(4) is convex in this constructed Riemannian

manifold. The key step would be the definition of the Riemannian metric in the feasible solutions set as $\mathcal{W} = \{ \{w^{(1)}, w^{(2)}, \dots, w^{(N)}\} | w^{(n)} \in \mathbb{R}^I, n = 1, 2, \dots, N \}$.

While the number of patients used in this work is comparable to a similar study (Wu et al., 2016), it's still relative small as compared to some other radiomic studies. Thus large steps in the ROC curves are observed due to the limited number of patients in both training and testing dataset. For the same reason, there is a possibility that overfitting may happen in the model training. However, our results indicate that for many models (e.g., LSTM and RF-WLCX using PET, and SVM-SFS, RF-WLCX and KSTM models using CT), the performance was comparable between training and testing datasets. The presented KSTM model performed fairly well on the test dataset, achieving AUCs of 0.74, 0.80 and 0.84 for PET, CT and fused PET-CT model, respectively. Furthermore, the results on the independent dataset demonstrate the KSTM model outperforms all other comparative methods investigated in this work. However, a larger independent patient cohort preferably including patients from other institutions is needed in order to validate the performance of KSTM.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was partly supported by the American Cancer Society (ACS-IRG-02-196), the National Institutes of Health (5P30CA142543) and the National Natural Science Foundation of China (NSFC, 11771456). The authors would like to thank Dr. Damiana Chiavolini for editing the manuscript.

References

- Abdi H, Williams LJ, 2010 Principal component analysis. Wiley interdisciplinary reviews: computational statistics 2, 433–459.
- Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D, 2014 Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nature communications 5.
- Ahmed A, 2011 Image denoising using Gabor filter banks, Computers & Informatics (ISCI), 2011 IEEE Symposium on IEEE, pp. 215–218.
- Biswas SK, Milanfar P, 2016 Linear Support Tensor Machine: Pedestrian Detection in Thermal Infrared Images. arXiv preprint arXiv:1609.07878.
- Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP, 2002 SMOTE: synthetic minority oversampling technique. Journal of artificial intelligence research 16, 321–357.
- Chi A, Liao Z, Nguyen NP, Xu J, Stea B, Komaki R, 2010 Systemic review of the patterns of failure following stereotactic body radiation therapy in early-stage non-small-cell lung cancer: clinical implications. Radiotherapy and Oncology 94, 1–11. [PubMed: 20074823]
- Chung AG, Shafiee MJ, Kumar D, Khalvati F, Haider MA, Wong A, 2015 Discovery Radiomics for Multi-Parametric MRI Prostate Cancer Detection. arXiv preprint arXiv:1509.00111.
- Ciampi F, de Hoop B, van Riel SJ, Chung K, Scholten ET, Oudkerk M, de Jong PA, Prokop M, van Ginneken B, 2015 Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box. Medical image analysis 26, 195–202. [PubMed: 26458112]
- Clarke K, Taremi M, Dahele M, Freeman M, Fung S, Franks K, Bezjak A, Brade A, Cho J, Hope A, 2012 Stereotactic body radiotherapy (SBRT) for non-small cell lung cancer (NSCLC): is FDGPET a predictor of outcome? Radiotherapy and Oncology 104, 62–66. [PubMed: 22682749]

- Cook GJ, Siddique M, Taylor BP, Yip C, Chicklore S, Goh V, 2014 Radiomics in PET: principles and applications. *Clinical and Translational Imaging* 2, 269–276.
- Cortes C, Vapnik V, 1995 Support-vector networks. *Machine learning* 20, 273–297.
- Davis LS, Johns SA, Aggarwal J, 1979 Texture analysis using generalized co-occurrence matrices. *IEEE Transactions on pattern analysis and machine intelligence*, 251–259. [PubMed: 21868856]
- De Lathauwer L, 1997 Signal processing based on multilinear algebra. Katholieke Universiteit Leuven.
- Ettinger DS, Akerley W, Bepler G, Blum MG, Chang A, Cheney RT, Chirieac LR, D'Amico TA, Demmy TL, Ganti AKP, 2010 Non-small cell lung cancer. *Journal of the national comprehensive cancer network* 8, 740–801. [PubMed: 20679538]
- Fletcher RH, Fletcher SW, Fletcher GS, 2012 *Clinical epidemiology: the essentials*. Lippincott Williams & Wilkins.
- Gillies RJ, Kinahan PE, Hricak H, 2015 Radiomics: images are more than pictures, they are data. *Radiology* 278, 563–577. [PubMed: 26579733]
- Ginsberg RJ, Rubinstein LV, Group LCS, 1995 Randomized trial of lobectomy versus limited resection for T1 N0 non-small cell lung cancer. *The Annals of Thoracic Surgery* 60, 615–623. [PubMed: 7677489]
- Hao Z, He L, Chen B, Yang X, 2013 A linear support higher-order tensor machine for classification. *IEEE Transactions on Image Processing* 22, 2911–2920. [PubMed: 23529095]
- Howlander N, N. A. Krapcho M, Miller D, Bishop K, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds). *SEER Cancer Statistics Review, 1975–2013*, National Cancer Institute Bethesda, MD, , 2016 http://seer.cancer.gov/csr/1975_2013/ based on November 2015 SEER data submission, posted to the SEER website, .
- Huang M, Yang W, Feng Q, Chen W, Initiative A.s.D.N., 2017 Longitudinal measurement and hierarchical classification framework for the prediction of Alzheimer's disease. *Scientific Reports* 7.
- Keogh E, Mueen A, 2011 Curse of dimensionality, *Encyclopedia of Machine Learning*. Springer, pp. 257–258.
- Khamis H, Zurakhov G, Azar V, Raz A, Friedman Z, Adam D, 2017 Automatic apical view classification of echocardiograms using a discriminative learning dictionary. *Medical Image Analysis* 36, 15–21. [PubMed: 27816858]
- Kohavi R, John GH, 1997 Wrappers for feature subset selection. *Artificial intelligence* 97, 273–324.
- Kolda TG, Bader BW, 2009 Tensor decompositions and applications. *SIAM review* 51, 455–500.
- Kononenko I, Simec E, Robnik-Šikonja M, 1997 Overcoming the myopia of inductive learning algorithms with RELIEFF. *Applied Intelligence* 7, 39–55.
- Kooi T, Litjens G, van Ginneken B, Gubern-Mérida A, Sánchez CI, Mann R, den Heeten A, Karssemeijer N, 2017 Large scale deep learning for computer aided detection of mammographic lesions. *Medical image analysis* 35, 303–312. [PubMed: 27497072]
- Krizhevsky A, Sutskever I, Hinton GE, 2012 Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, pp. 1097–1105.
- Kumar D, Shafiee MJ, Chung AG, Khalvati F, Haider MA, Wong A, 2015 Discovery radiomics for computed tomography cancer detection. *arXiv preprint arXiv:1509.00117*.
- Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, Forster K, Aerts HJ, Dekker A, Fenstermacher D, 2012 Radiomics: the process and the challenges. *Magnetic resonance imaging* 30, 1234–1248. [PubMed: 22898692]
- Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, Zegers CM, Gillies R, Boellard R, Dekker A, 2012 Radiomics: extracting more information from medical images using advanced feature analysis. *European Journal of Cancer* 48, 441–446. [PubMed: 22257792]
- Li S-L, Li C, Liou Y-C, Yao J-C, 2009 Existence of solutions for variational inequalities on Riemannian manifolds. *Nonlinear Analysis: Theory, Methods & Applications* 71, 5695–5706.

- Lian C, Ruan S, Dencœux T, Jardin F, Vera P, 2016 Selecting radiomic features from FDG-PET images for cancer treatment outcome prediction. *Medical image analysis* 32, 257–268. [PubMed: 27236221]
- Liu H-W, Gabos Z, Ghosh S, Roberts B, Lau H, Kerba M, 2015 Outcomes in stage I non-small cell lung cancer following the introduction of stereotactic body radiotherapy in Alberta–A population-based study. *Radiotherapy and Oncology* 117, 71–76. [PubMed: 26349589]
- Liu M, Yu N, Li W, 2010 Camera model identification for JPEG images via tensor analysis, *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, 2010 Sixth International Conference on IEEE, pp. 462–465.
- Liu M, Zhang J, Yap P-T, Shen D, 2017 View-aligned hypergraph learning for Alzheimer’s disease diagnosis with incomplete multi-modality data. *Medical image analysis* 36, 123–134. [PubMed: 27898305]
- Lovinfosse P, Janvary ZL, Coucke P, Jodogne S, Bernard C, Hatt M, Visvikis D, Jansen N, Duysinx B, Hustinx R, 2016 FDG PET/CT texture analysis for predicting the outcome of lung cancer treated by stereotactic body radiation therapy. *European journal of nuclear medicine and molecular imaging* 43, 1453–1460. [PubMed: 26830299]
- Madabhushi A, Lee G, 2016 Image analysis and machine learning in digital pathology: Challenges and opportunities. *Medical Image Analysis* 33, 170–175. [PubMed: 27423409]
- Magnin B, Mesrob L, Kinkingnéhun S, Péligrini-Issac M, Colliot O, Sarazin M, Dubois B, Lehericy S, Benali H, 2009 Support vector machine-based classification of Alzheimer’s disease from whole-brain anatomical MRI. *Neuroradiology* 51, 73–83. [PubMed: 18846369]
- Martini N, Bains MS, Burt ME, Zakowski MF, McCormack P, Rusch VW, Ginsberg RJ, 1995 Incidence of local recurrence and second primary tumors in resected stage I lung cancer. *The Journal of thoracic and cardiovascular surgery* 109, 120–129. [PubMed: 7815787]
- Miller KD, Siegel RL, Lin CC, Mariotto AB, Kramer JL, Rowland JH, Stein KD, Alteri R, Jemal A, 2016 Cancer treatment and survivorship statistics, 2016. *CA: a cancer journal for clinicians* 66, 271–289. [PubMed: 27253694]
- Miotto R, Wang F, Wang S, Jiang X, Dudley JT, 2017 Deep learning for healthcare: review, opportunities and challenges. *Briefings in Bioinformatics*, bbx044.
- Namburete AI, Stebbing RV, Kemp B, Yaqub M, Papageorgiou AT, Noble JA, 2015 Learning-based prediction of gestational age from ultrasound images of the fetal brain. *Medical image analysis* 21, 72–86. [PubMed: 25624045]
- Nath SK, Sandhu AP, Kim D, Bharne A, Nobiensky PD, Lawson JD, Fuster M, Bazhenova L, Song WY, Mundt AJ, 2011 Locoregional and distant failure following image-guided stereotactic body radiation for early-stage primary lung cancer. *Radiotherapy and Oncology* 99, 12–17. [PubMed: 21429608]
- Nikolaev A, Benda R, Shang C, Schramm A, Watson S, Kasper M, Williams T, 2016 Significance of Tumor Cell Histology for Local Control of Early-Stage Non-Small Cell Lung Cancer Treated With Stereotactic Body Radiation Therapy: A Stratified Retrospective Analysis. *International Journal of Radiation Oncology Biology Physics* 96, E424.
- Otsu N, 1979 A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics* 9, 62–66.
- Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJ, 2015 Machine learning methods for quantitative radiomic biomarkers. *Scientific reports* 5, 13087. [PubMed: 26278466]
- Raudys SJ, Jain AK, 1991 Small sample size effects in statistical pattern recognition: Recommendations for practitioners. *IEEE Transactions on pattern analysis and machine intelligence* 13, 252–264.
- Roth HR, Lu L, Liu J, Yao J, Seff A, Cherry K, Kim L, Summers RM, 2016 Improving Computer-Aided Detection Using Convolutional Neural Networks and Random View Aggregation. *IEEE transactions on medical imaging* 35, 1170–1181. [PubMed: 26441412]
- Sathyamurthy N, Raff L, 1975 Quasiclassical trajectory studies using 3D spline interpolation of abinitio surfaces. *The Journal of Chemical Physics* 63, 464–473.

- Senthi S, Lagerwaard FJ, Haasbeek CJ, Slotman BJ, Senan S, 2012 Patterns of disease recurrence after stereotactic ablative radiotherapy for early stage non-small-cell lung cancer: a retrospective analysis. *The lancet oncology* 13, 802–809. [PubMed: 22727222]
- Shashua A, Levin A, 2001 Linear image coding for regression and classification using the tensor-rank principle, *Computer Vision and Pattern Recognition*, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on IEEE, pp. I–I.
- Shen W, Zhou M, Yang F, Yang C, Tian J, 2015 Multi-scale convolutional neural networks for lung nodule classification, *International Conference on Information Processing in Medical Imaging* Springer, pp. 588–599.
- Shin H-C, Roth HR, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D, Summers RM, 2016 Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging* 35, 1285–1298. [PubMed: 26886976]
- Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, Liang J, 2016 Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE transactions on medical imaging* 35, 1299–1312. [PubMed: 26978662]
- Tan S, Kligerman S, Chen W, Lu M, Kim G, Feigenberg S, D'Souza WD, Suntharalingam M, Lu W, 2013 Spatial-temporal [18 F] FDG-PET features for predicting pathologic response of esophageal cancer to neoadjuvant chemoradiation therapy. *International Journal of Radiation Oncology Biology Physics* 85, 1375–1382.
- Tao D, Li X, Hu W, Maybank S, Wu X, 2005 Supervised tensor learning, *Fifth IEEE International Conference on Data Mining (ICDM'05) IEEE*, p. pp. 8.
- Tao D, Li X, Wu X, Hu W, Maybank SJ, 2007 Supervised tensor learning. *Knowledge and Information Systems* 13, 1–42.
- Team NLSTR, 2013 Results of initial low-dose computed tomographic screening for lung cancer. *N Engl J Med* 2013, 1980–1991.
- Timmerman R, Hu C, Michalski J, Straube W, Galvin J, Johnstone D, Bradley J, Barriger R, Bezjak A, Videtic G, 2014 Long-term results of RTOG 0236: a phase II trial of stereotactic body radiation therapy (SBRT) in the treatment of patients with medically inoperable stage I non-small cell lung cancer. *International Journal of Radiation Oncology Biology Physics* 90, S30.
- Timmerman R, Paulus R, Galvin J, Michalski J, Straube W, Bradley J, Fakiris A, Bezjak A, Videtic G, Johnstone D, 2010 Stereotactic body radiation therapy for inoperable early stage lung cancer. *Jama* 303, 1070–1076. [PubMed: 20233825]
- Vallièrès M, Freeman C, Skamene S, El Naqa I, 2015 A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Physics in medicine and biology* 60, 5471–5496. [PubMed: 26119045]
- Vansteenkiste JF, Stroobants SG, Dupont PJ, De Leyn PR, Verbeken EK, Deneffe GJ, Mortelmans LA, Demedts MG, 1999 Prognostic importance of the standardized uptake value on 18Ffluoro-2-deoxy-glucose–positron emission tomography scan in non–small-cell lung cancer: an analysis of 125 cases. *Journal of Clinical Oncology* 17, 3201–3206. [PubMed: 10506619]
- Wimmer G, Tamaki T, Tischendorf JJ, Häfner M, Yoshida S, Tanaka S, Uhl A, 2016 Directional wavelet based features for colonic polyp classification. *Medical image analysis* 31, 16–36. [PubMed: 26948110]
- World Health Organization, 2017 <http://www.who.int/mediacentre/factsheets/fs297/en/index.html>.
- Wu J, Aguilera T, Shultz D, Gudur M, Rubin DL, Loo BW, Jr, Diehn M, Li R, 2016 Early-Stage Non–Small Cell Lung Cancer: Quantitative Imaging Characteristics of 18F Fluorodeoxyglucose PET/CT Allow Prediction of Distant Metastasis. *Radiology* 281, 270–278. [PubMed: 27046074]
- Zhang L, Zhang L, Tao D, Huang X, 2011 A multifeature tensor for remote-sensing target recognition. *IEEE Geoscience and Remote Sensing Letters* 8, 374–378.
- Zhou Zhiguo, Z. Z-J, Hao Hongxia, Li Shulong, Chen Xi, Zhang You, Folkert Michael, Wang Jing, 2017 Constructing multi-modality and multi-classifier radiomics predictive models through reliable classifier fusion. *arXiv:1710.01614*.

- Zhou Z-G, Liu F, Jiao L-C, Li L-L, Wang X-D, Gou S-P, Wang S, 2013 Object information based interactive segmentation for fatty tissue extraction. *Computers in biology and medicine* 43, 1462–1470. [PubMed: 24034738]
- Zhou Z, Folkert M, Cannon N, Iyengar P, Westover K, Zhang Y, Choy H, Timmerman R, Yan J, Xie X-J, Jiang S, Wang J, 2016 Predicting distant failure in early stage NSCLC treated with SBRT using clinical parameters. *Radiotherapy and Oncology* 119, 501–504. [PubMed: 27156652]
- Zhou Z, Folkert M, Iyengar P, Westover K, Zhang Y, Choy H, Timmerman R, Jiang S, Wang J, 2017 Multi-objective radiomics model for predicting distant failure in lung SBRT. *Physics in medicine and biology* 62, 4460. [PubMed: 28480871]
- Zuluaga MA, Burgos N, Mendelson AF, Taylor AM, Ourselin S, 2015 Voxelwise atlas rating for computer assisted diagnosis: Application to congenital heart diseases of the great arteries. *Medical image analysis* 26, 185–194. [PubMed: 26433929]

Highlights:

- A kernelled support tensor machine (KSTM) is proposed to assess SBRT for NSCLC patients.
- The KSTM based model uses native tumor imaging as input.
- The KSTM trains the classification hyper-plane in an adaptive feature tensor space.
- An iterative procedure with the proof of convergence for KSTM is developed.
- The KSTM achieved the highest prediction results among the seven investigated methods.

Fig. 1.

The four constructed tumor tensors (bottom row) and their corresponding 2D patches of the tumor from one of slice (upper row): (a) and (b) for PET imaging; (c) and (d) for CT imaging; (a) and (c) show the same patient without distant failure; (b) and (d) show the same patient with distant failure.

Fig. 2.

The mode product between a 3D tensor and three vectors.

Fig. 3.
The function value of the left formula in (13) with respect to different iterations: (a) for PET tumor imaging with SUV and (b) for original CT tumor imaging.

Fig. 4.
Workflow of the KSTM-based classifier.

Fig. 5. Performance of the STM-based classifier with 3D spline interpolation and 3D Gabor filtering: a) performance for PET imaging; b) performance for CT imaging.

Fig. 6. ROC curves for the STM-based classifier with and without 3D spline interpolation and 3D Gabor filtering: (a) PET imaging and (b) CT imaging.

Fig. 7. ROC curves for seven compared predictive algorithms for PET imaging: (a) 10-fold cross validation and (b) independent validation.

Fig. 8. ROC curves for seven compared predictive algorithms for CT imaging: (a) 10-fold cross validation and (b) independent validation.

Fig. 9. ROC curves of the KSTM-based classifier for PET and CT, and the multi-modality fusion of PET and CT based on RCF.

Table 1:

The KSTM iterative algorithm

Input: A set of tensor samples $\left\{X_m \in R^{J_1 \times J_2 \times \dots \times J_N}, y_m \in \left\{-1, 1\right\}\right\}_{m=1}^M$, threshold parameters ϵ .

Output: $w^{(l)} \in R^{J_l} (l = 1, 2, \dots, N, l \neq n)$, $\alpha_m^n \in R^{I_n} (m = 1, 2, \dots, M)$ and $\theta^{(n)}$.

Step 1: Set $w^{(l)} \in R^{J_l}$ equal to unit vector with same elements $w_0^{(l)}, l = 1, 2, \dots, N$, and the initial projective mappings P_l in (5)–(7) are defined as

$$P_l(w_0^{(l)}) = w_0^{(l)}. \quad (8)$$

Step 2: Fix $w^{(l)} (l = 1, 2, \dots, N, l \neq n)$ to be the known weight vectors.

Step 3: For $n = 1, 2, \dots, N$, set $Y_i^{(n)} = X_i \prod_{1 \leq l \leq N, l \neq n} P_l(w^{(l)})$, $i = 1, 2, \dots, M$.

Step 4: Assume $w^{(n)} = \sum_{i=1}^M \alpha_i^{(n)} y_i \varnothing_n(Y_i^{(n)})$, and solve the dual problem of the quadratic program problem (5)–(7) by SVM solver with kernel to obtain $\alpha_i^{(n)}, i = 1, 2, \dots, M$ as follows:

$$\operatorname{argmax} J(\alpha_i^{(n)} (i = 1, 2, \dots, M)) = \sum_{i=1}^M \alpha_i^{(n)} - \frac{1}{2} \sum_{i,j=1}^M \alpha_i^{(n)} \alpha_j^{(n)} y_i y_j K(Y_i^{(n)}, Y_j^{(n)}), \quad (9)$$

subjected to

$$0 < \alpha_i^{(n)} \leq C, i = 1, 2, \dots, M, \quad (10)$$

$$\sum_{i=1}^M \alpha_i^{(n)} y_i = 0, \quad (11)$$

where K is a kernel function satisfying $K(Y_i^{(n)}, Y_j^{(n)}) = \langle \varnothing_n(Y_i^{(n)}), \varnothing_n(Y_j^{(n)}) \rangle$.

Step 5: Set

$$P_l(w^{(l)}) = \sum_{i=1}^M \alpha_i^{(l)} y_i Y_i^{(l)}, l = n. \quad (12)$$

Step 6: Let $n \leftarrow n+1$ if $n < N-1$ and $n \leftarrow 1$ if $n = N$.

Repeat Step 3-Step6 until convergence i.e. satisfying (13).

Step 7: If

$$\left| \sum_{l=1}^N (\langle P_l(w_t^{(l)}), P_l(w_{t-1}^{(l)}) \rangle - \|P_l(w_t^{(l)})\|_F^{-2} - 1) \right| < \varepsilon, \quad (13)$$

set $w^{(l)} = P_l(w_t^{(l)}) \in R^J$ ($l = 1, 2, \dots, N$, $l \neq n$) and output them, then go Step 8. Otherwise, go to Step 3 to Step 7. Here t represents the t th iteration.

Step 8: Set $Y_i^{(n)} = X_i \prod_{1 \leq l \leq N, l \neq n} w^{(l)}$, $i = 1, 2, \dots, M$ and solve the optimization problem (9)–(11) by SVM solver to obtain the final $\alpha_i^{(n)}$, $i = 1, 2, \dots, M$ and $b^{(n)}$. Then output $\alpha_i^{(n)}$, $i = 1, 2, \dots, M$, $b^{(n)}$, and stop.

Table 2:

Performance of different predictive models for PET.

Methods	10-fold cross validation				Independent testing			
	AUC	ACC (%)	SEN (%)	SPE (%)	AUC	ACC (%)	SEN (%)	SPE
SVM	0.72±0.03	73.51±3.96	64.44±6.33	76.27±6.23	0.58	60.61	42.86	65.38
SVM-PCA	0.70±0.06	72.47±3.94	58.89±16.94	76.61±9.08	0.61	66.67	71.43	65.38
LSTM	0.74±0.03	73.77±3.10	62.22±11.39	77.29±4.58	0.72	72.73	42.86	80.77
SVM-ReliefF	0.79±0.01	73.25±4.91	76.67±7.24	72.20±6.74	0.73	69.70	71.43	69.23
SVM-SFS	0.77±0.03	71.43±3.67	72.22±7.86	71.19±7.19	0.69	72.73	71.43	73.08
RF-WLCX	0.65±0.06	66.49±1.93	55.56±8.78	69.83±3.26	0.64	63.64	71.43	61.54
KSTM	0.81±0.02	76.10±3.51	78.89±5.69	75.25±7.45	0.74	75.76	71.43	76.92

Table 3:

Performance of different predictive models for CT.

Methods	10-fold cross validation				Independent testing			
	AUC	ACC (%)	SEN (%)	SPE (%)	AUC	ACC (%)	SEN (%)	SPE (%)
SVM	0.75±0.05	77.92±3.67	71.11±12.67	80.00±7.89	0.60	57.58	42.86	61.54
SVM-PCA	0.75±0.05	76.36±2.13	67.78±6.09	78.98±4.42	0.69	72.73	57.14	76.92
LSTM	0.69±0.08	64.94±4.11	52.22±16.48	68.81±8.09	0.62	66.67	57.14	69.23
SVM-ReliefF	0.76±0.04	70.65±3.51	73.33±4.65	69.83±4.65	0.67	72.73	71.43	73.08
SVM-SFS	0.74±0.03	75.84±2.69	64.44±6.33	79.32±4.70	0.71	66.67	71.43	65.38
RF-WLCX	0.63±0.01	65.45±2.17	52.22±7.45	69.46±3.17	0.64	69.70	57.14	73.08
KSTM	0.81±0.03	78.44±3.63	77.78±7.86	78.64±6.95	0.80	75.76	71.43	76.92

Table 4:

The p-values in the unpaired T-test between the performances of the KSTM based classifier and each competing method.

10-fold cross validation		
Method	PET	CT
SVM	< 0.0001	0.0434
SVM-PCA	0.0028	0.0496
LSTM	0.0011	0.0016
SVM-ReliefF	0.0487	0.0294
SVM-SFS	0.0234	0.0032
RF-WLCX	< 0.0001	< 0.0001

Table 5:

The results of the multi-modality fusion of PET and CT based on RCF.

Method	AUC	ACC (%)	SEN (%)	SPE (%)
RCF-PET-CT	0.84	84.85	71.43	88.46