Computerized Medical Imaging and Graphics xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

Transfer learning on fused multiparametric MR images for classifying histopathological subtypes of rhabdomyosarcoma

Imon Banerjee*, Alexis Crawley, Mythili Bhethanabotla, Heike E Daldrup-Link, Daniel L. Rubin

Department of Radiology, Stanford University School of Medicine, Stanford, CA, United States of America

ARTICLE INFO

Article history:
Received 31 January 2017
Received in revised form 30 April 2017
Accepted 4 May 2017

Keywords:
Rhabdomyosarcoma
Computer aided diagnosis
Image fusion
Transfer learning
Deep neural networks

ABSTRACT

This paper presents a deep-learning-based CADx for the differential diagnosis of embryonal (ERMS) and alveolar (ARMS) subtypes of rhabdomysarcoma (RMS) solely by analyzing multiparametric MR images. We formulated an automated pipeline that creates a comprehensive representation of tumor by performing a fusion of diffusion-weighted MR scans (DWI) and gadolinium chelate-enhanced T1−weighted MR scans (MRI). Finally, we adapted transfer learning approach where a pre-trained deep convolutional neural network has been fine-tuned based on the fused images for performing classification of the two RMS subtypes. We achieved 85% cross validation prediction accuracy from the fine-tuned deep CNN model. Our system can be exploited to provide a fast, efficient and reproducible diagnosis of RMS subtypes with less human interaction. The framework offers an efficient integration between advanced image processing methods and cutting-edge deep learning techniques which can be extended to deal with other clinical domains that involve multimodal imaging for disease diagnosis.

1. Introduction

Rhabdomyosarcoma (RMS) represents the most common extracranial solid malignancy in children and adolescents with an age of less than 20 years (Ognjanovic et al., 2009). The majority of RMS is of embryonal (ERMS) and alveolar (ARMS) subtypes. Patient outcomes vary considerably, with 5 years survival rates between 35% and 95% depending on the type of RMS involved, tumor grade, tumor stage and patient age, among other factors (Malempati and Hawkins, 2012). Most ARMS are more aggressive than ERMS and require more intense treatment. A diagnosis of the histopathological subtype is critical for effective personalized treatment and survival. In the clinic, RMS subtypes are classified based on specific morphological and genetic characteristics, obtained from the biopsy specimens.

Medical imaging can contribute to the classification of RMS subtypes based on tumor location, but traditional imaging findings are non-specific. A few clinical studies (Baum et al., 2011; Brenner et al., 2004) have shown that the extent of 18F-FDG uptake (representing tumor metabolism) on PET images and the degree of restricted diffusion on MR images (representing tumor cell density) can be linked to prognostic information. While these studies showed some correlation between tumor metabolism/diffusion and patient survival, the degree or distribution of 18F-FDG uptake or diffusion restriction in the tumor tissue is not yet established as an identifier for differentiating ERMS and ARMS. New imaging signs to the categorization of RMS tumors into high and low risk groups could potentially improve assignment of treatment options and outcomes. To the best of our knowledge, no computerized diagnosis system exists that can classify ERMS from ARMS by analyzing only the organ-level scans (e.g. MRI, DWI, PET).

The purpose of our study is to differentiate ARMS and ERMS by analyzing a fusion of diffusion weighted MR and T1 weighted contrast enhanced MR images with less manual intervention. The extraction of effective image features for the differentiation of RMS subtypes is the most crucial component of this study. However, it is an extremely complicated task due to the need of hand crafted descriptor design/selection which requires much manual effort and a deep investigation of the data. We hypothesize that deep convolutional neural network (CNN) based RMS classifier that learns automatically informative features from the fused multiparametric MR images, can provide an effective and convenient solution for the differentiation of ERMS and ARMS subtypes. We also believe that transfer learning approach can be more suitable for our study. Mainly due to the fact that the current dataset is restricted in size, and therefore inappropriate to train a deep-

* Corresponding author.
E-mail address: imonb@stanford.edu (I. Banerjee).

http://dx.doi.org/10.1016/j.compmedimag.2017.05.002
0895-6111/© 2017 Elsevier Ltd. All rights reserved.

CNN from scratch. But, the CNNs comprehensively trained on the large scale well-annotated ImageNet dataset (contains 1.2 million images labeled with 1000 categories) may be transferred and fine-tuned on the small medical dataset for performing the RMS classification, regardless the disparity between natural images and fused MR scan images.

We developed a radiomics framework that classifies ARMS from ERMS tumors by exploiting a tight integration of advanced image processing methods and cutting-edge deep learning techniques. First, we independently segmented the tumor from two multiparametric MR images (contrast enhanced T1 weighted MR image and diffusion weighted MR image) using a completely automatic segmentation pipeline. Afterwards, we registered the segmented tumor images using a sophisticated non-rigid registration technique and generated the fused RGB color images from the registered data. We applied standard data augmentation technique to obtain a sufficiently large training dataset to train the CNN. Finally, the fused RGB images were used to train a deep neural network in which we adapted a transfer learning approach with AlexNet model for the RMS classification task. During the training phase, we fine-tuned the ImageNet pre-trained AlexNet model and achieved 85% cross validation classification accuracy. The following sections (2, 3, 4, 5) give further details on the dataset, procedure and explain how the segmentation, registration, data augmentation, and deep learning aspects were practiced focusing our scientific contributions in each step.

2. Related works

Tumor classification by integrating texture analysis of medical images and standard machine learning techniques, is a common approach in literature. In Othman et al. (2011) authors performed classification of brain tumor using Daubechies (db4) wavelet texture analysis and Support Vector Machine (SVM) and 65% accuracy was obtained, where, only 39 images were successfully classified from 60 images. It was concluded that classification using Support Vector Machine resulted in a limited precision, since it cannot work accurately for a large data due to training complexity. In Othman and Basri (2011), a Probabilistic Neural Network (PNN) for tumor classification was proposed to classify brain tumor using Principal Component Analysis for feature extraction and PNN for classification. They concluded that PNN is a promising tool for brain tumor classification, based on its fast speed and its accuracy which ranges from 73 to 100% for spread values (smoothing factor) from 1 to 3. Classification of brain MRI using the LH and HL wavelet transform sub-bands was performed in Lahmiri and Boukadoum (2011) that shows that feature extraction from the LH (Low-High) and HL (High-Low) sub-bands using first order statistics has higher performance than features from LL (Low-low) bands. A few studies (Mayerhoefer et al., 2008; Juntu et al., 2010) showed that texture analysis can also be informative for discrimination between benign and malignant soft-tissue sarcomas in MRI images. However, such standard machine learning techniques need very specific feature extractors for each type of tumor classification task and this requires much manual data analysis, and it becomes difficult to extend the approaches for a new dataset.

In the last few years, deep convolutional neural networks (Deep-CNNs) that try to learn high level features from the given data, has been successfully applied to a wide range of applications, including natural language processing, image classification, semantic tagging (LeCun et al., 2015). Deep-CNN is reducing the task of making new feature extractor for each type of data (speech, image, etc.). Recently, the Deep-CNNs have also been introduced to the medical domain with promising results in various areas, like organ segmentations and detection, image standard plane selection, computerized diagnosis and prognosis, etc. (Greenspan et al., 2016). The Deep-CNN could potentially change the design paradigm of the computerized diagnosis and prognosis framework due to several advantages over the standard machine learning. First, deep learning can directly uncover features from the training data, and hence the effort of explicit elaboration on feature extraction can be significantly alleviated (Bengio et al., 2007). The neuron-crafted features may compensate and even surpass the discriminative power of the conventional feature extraction methods. Second, feature interaction and hierarchy can be exploited jointly within the intrinsic deep architecture of a neural network (Lee et al., 2011). Consequently, the feature selection process will be significantly simplified. Third, the three steps of feature extraction, selection and supervised classification can be realized within the optimization of the same deep architecture (Krizhevsky et al., 2012). With such a design, the performance can be tuned more easily in a systematic fashion.

Recently, ImageNet pre-trained CNNs have been used for chest pathology identification and detection in X-ray and CT modalities (van Ginneken et al., 2015; Bar et al., 4140; Ciompi et al., 2015) and have yielded the best performance results. However, the fine-tuning of a pre-trained CNN model on multimodal RMS tumor image datasets has not yet been exploited. In this work, we propose to use the transfer leaning approach where we use pre-trained AlexNet Deep-CNN for classifying aggressive ARMS from less aggressive ERMS brain tumor by using fusion of T1 weighted contrast enhance MR and diffusion weighted MR images.

3. Materials and method

Fig. 1 illustrates a simplified schematic diagram of the proposed pipeline where we show how an unseen dataset of multiparametric MR images will be automatically analyzed and classified within our proposed framework. In the following sub-sections, we detail each of the core components including description of the dataset used in the study and the model training methodology.

3.1. Dataset

In a retrospective, *Institutional Review Board* (IRB) approved study, we evaluated diffusion-weighted MR scans and contrast enhanced T1-weighted MR scans of 21 children and adolescents (age 1–20 years) with newly diagnosed intermediate-risk ARMS ($n = 6$) and ERMS ($n = 15$) in the head and neck. The cohort included 7 girls and 14 boys with ages between 1–20 years (10.04 ± 5.42 SD). To our knowledge, this is the largest multi-institutional cohort of children with RMS imaged with these two modalities to date. The imaging work-flow consisted of contrast enhanced T1 weighted MR scans (MRI) and diffusion weighted MR scans (DWI) before initiation of chemotherapy. The scans are acquired in the axial plane for all patients. An intravenous infusion of gadolinium chelate was used as contrasting agent. The dose of the contrast agent was decided based on the infant's weight to get a similar distribution in the patient's blood. The MR scans' slice thickness was 3–5 mm and in-plane resolution range was 0.4–0.5 mm^2. The diffusion weighted MR scans' slice thickness was 2.5–5 mm and in plane resolution range was 1–1.2 mm^2. The shorter b-value in the range of 500–700 s/mm has been used.

Tumor was outlined as 2D region of interest (ROI) on single slice of each scan by the certified radiologists via an interactive web-based software-epad (Rubin et al., 2014). In Fig. 2, we show an epad snapshot of how the annotation has been done. Through visual inspection of each scan volume, the radiologist also determined the slice range where the tumor is visible. For the current dataset, there were between 5–8 slices. For each patient data, RMS

Please cite this article in press as: Banerjee, I., et al., Transfer learning on fused multiparametric MR images for classifying histopathological subtypes of rhabdomyosarcoma. Comput Med Imaging Graph (2017), http://dx.doi.org/10.1016/j.compmedimag.2017.05.002

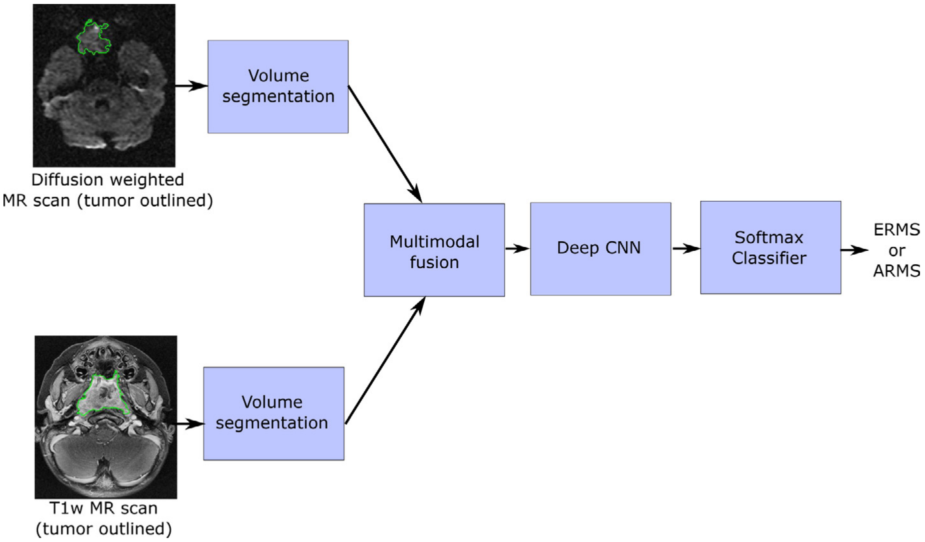

Fig. 1. Simple schematic diagram of RMS classification in our proposed method.

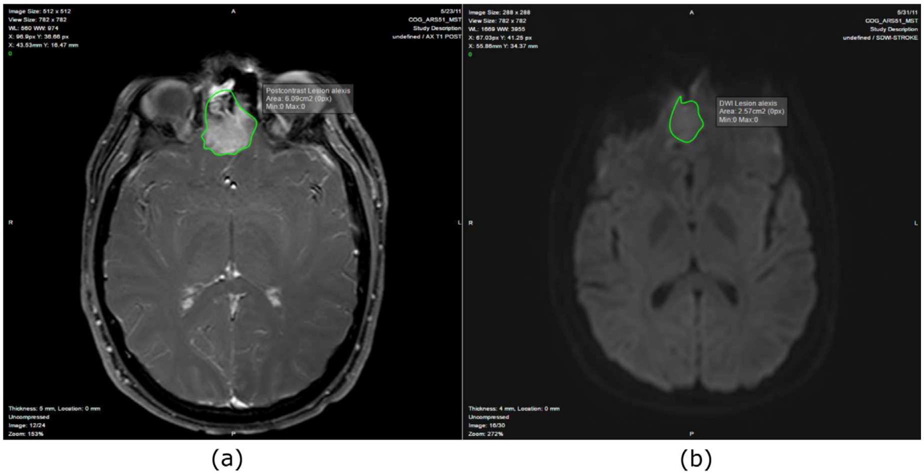

Fig. 2. Example of annotated images on epad software. The MR image (on the left) and DWI (on the right) for a certain patient with the hand drawn annotations by a certified radiologist.

subtype has been determined by biopsy result, which served as the ground truth.

3.2. *Segmentation of tumor from MRI and DWI: propagation of level set*

For segmenting the tumor from diffusion weighted and T1 weighted MR scan volumes, we choose to use an extended version level set segmentation (Li et al., 2011) which is a non-parametric deformable model based segmentation method. The level set method can handle topological changes during curve evolution and is able to identify the object boundary by handling the challenging characteristics of medical images, namely shape variations, image noise, intensity heterogeneities, and discontinuous object boundaries. In 2D, the object boundary in level set can be represented by a closed curve: $\tau = (x, y), \phi(x, y) = 0$, where ϕ is the level set function. Given an initial ϕ at time $t = 0$ and its motion over time, it is possible to know $\phi(x, y, t)$ at any time t by evolving the initial $\phi(x, y, t = 0)$ over time. Vese and Chan (2002) proposed a global framework of piecewise constant level set model which is further expanded in Hoogi et al. (2016) by adding an Adaptive Local Window approach that is particularly relevant for segmenting heterogeneous lesion from CT and MR images. In this approach, the local window size is re-estimated at each point of the image (x, y) by an iterative process that considers the object scale, local and global texture statistics, and minimization of the cost function.

The initialization of the levelset is an important measure of the adaptive window level set segmentation performance, otherwise, it may converge to a local minimum and fail to capture the accurate boundary of the object. The most common techniques for the contour initialization are (1) manual selection of initial points (Ardon and Cohen, 2006); (2) analysis of the external force field (He et al., 2006); (3) naive geometric models such as a circle in 2-D or sphere in 3-D; and, (4) learnt shape priors, where a statistical shape model is estimated (Freedman and Zhang, 2005). However, methods based on shape priors may be restrictive in applications involving highly variable shapes.

In this paper, we propose a simple but efficient initialization technique that not only provides an accurate initialization of the level set, but also provides an easy extension of the 2D level set segmentation method (Hoogi et al., 2016) for segmenting 3D volume of MRI and DWI. Given the tumor boundary $ROI_i(x, y)$ identified by a radiologist on a single slice (S_i) of the MRI and DWI scans and the slice range where the tumor is visible, we segment the whole

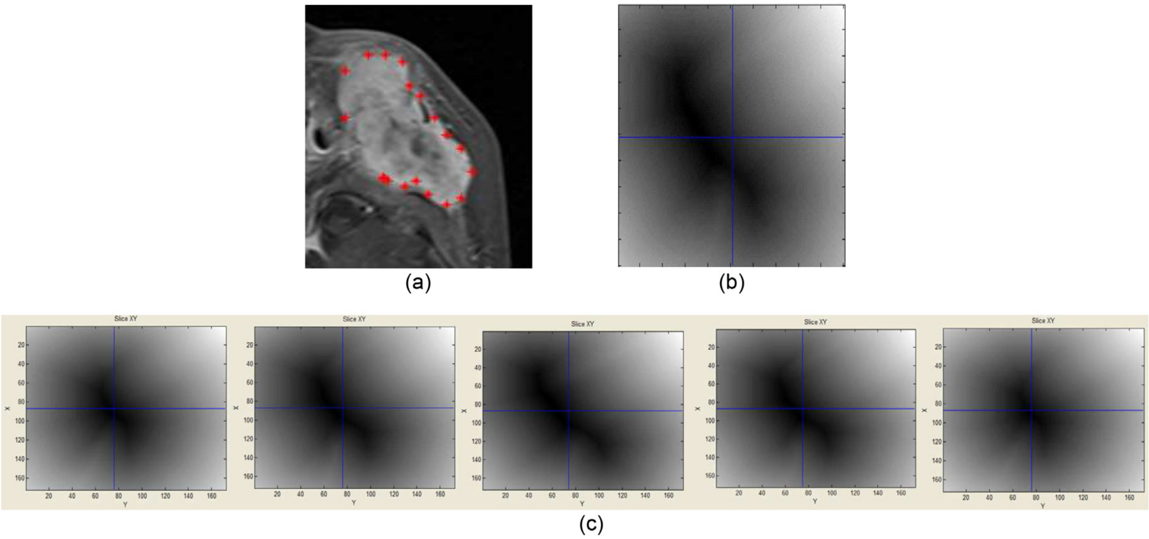

Fig. 3. Erosion operation on SDF image: (a) tumor ROI_i manually drawn on slice S_i; (b) SDF image computed from ROI_i; (c) result of erosion operation.

volume by propagating the object boundary with a new initialization technique. First, we define the $ROI(x, y)$ as zero level set and initialize ϕ by computing signed distance image (SDF) from the ROI_i as: $\phi(x, y, t=0)=\pm d$, where (x, y) are the pixel positions in the slice, the zero-level set which is obtained by using the $ROI(x, y)$ coordinates, and d is the minimum distance between position of the zero-level set. In order to segment the immediate upper and lower slice of (S_i), we apply a morphological operation on the initial $ROI_i(x, y)$ where the ROI_i is eroded by a circular structuring element $s_i(x, y)$ as: $ROI_(i+1)=ROI_i(x, y)\Theta s_i(x, y)$. The radius of the structuring element is decided empirically by evaluating the scanned slice thickness. The erosion operation basically retains the original shape while making it smaller in size. The main hypothesis behind the erosion operation is that the radiologist outlined the tumor on the slice where maximum tumor area is visible, and therefore on the upper and lower slices the tumor shape should be similar but of smaller size. We propagate the ROI_{i+1} on the immediate upper (S_{i+1}) and lower (S_{i-1}) slice of volume. In Fig. 3, we show the SDF images when the erosion performed on 5 different consecutive slices of the MR volume.

The foregoing steps give us more accurate initialization for performing level set segmentation on each slice of the scanned volumes since the tumor shape and tumor location are more closely approximated with the propagation technique. We execute the level set segmentation for each slice independently throughout the range given by the radiologist where the tumor is visible. Once the targeted slice is segmented, we propagate the current ROI onto the immediate next slice which is not segmented and repeat the same steps for the whole slice range.

With this initialization methodology, with the tumor outlined on a single slice, we segmented both MRI and DWI volumes using the 2D adaptive window level set segmentation method proposed in Hoogi et al. (2016). In Fig. 4, we show the MRI and DWI volume segmentation results for a patient where the whole volume has been segmented using the new initialization and the adaptive window level set segmentation method. The whole segmentation pipeline is implemented as a Matlab function which only needs the scanned volume, slice range and tumor outlined on a slice as input, and produces segmentation of the tumor from the whole scanned volume. Other than the manual marking on a single slice, our segmentation of MRI and DWI volume is a completely automated process.

3.3. Multi-modal image fusion: non-rigid demon algorithm

Once the MRI and DWI images are segmented for each patient, the MRI and DWI volumes are automatically cropped based on the bounding box of the segmented volume and re-sampled to have equal size in both modalities. Next, we generate fused RGB images (16 bit per channel) by registering the MRI and DWI volumes. For registration, we apply the diffeomorphic log demons 3D image registration algorithm (Vercauteren et al., 2009) that overlays two equal sized volumes of different modalities in a computationally efficient way. We choose to use the diffeomorphic demons algorithm since it is able to register the 3D multimodal images by using the image similarity criterion, the mean squared error, with a smooth invertible transformation. The diffeomorphic log demons algorithm combines a Lie group framework on diffeomorphisms and an optimization procedure for Lie groups to provide non-parametric diffeomorphic transformations of the entire displacement field.

In the registration, we consider the cropped MR volume as the fixed image F(.) and the cropped DWI volume as the moving image M(.) and, due to difference in the spatial resolution, F(.) and M(.) can be of different size (see Section 3.1). To generate equal sized volumes for registration, we use trilinear interpolation technique which is the 3D extension of linear interpolation that approximates the value of an intermediate point (x, y, z) within the local axial rectangular prism using data on the lattice points. We re-sample F(.) and M(.) to spatial resolution of 227 × 227 × 7. At this point, the image registration is treated as an optimization problem that aims at finding the diffeomorphic transformations of each voxel of M(.) to get a reasonable alignment to F(.).

To leverage the CNN architectures designed for color images (3 channels) and to transfer CNN parameters pre-trained on ImageNet, we generate fused RGB image by overlapping fixed image with registered image using the traditional alfa channel compostsing technique (Porter and Duff, 1984). It combines translucent color of foreground registered DWI with the background MRI as convex combination and creates the transparency effect. The degree of the foreground color's translucency may range from completely transparent (value 1) to completely opaque (value 0). We empirically select 0.4 transparency value which will allow looking through the foreground DWI and detect the MR background image of the patient. In Fig. 5, we show the registration outcome for a single slice of the fixed image and the corresponding RGB registered image.

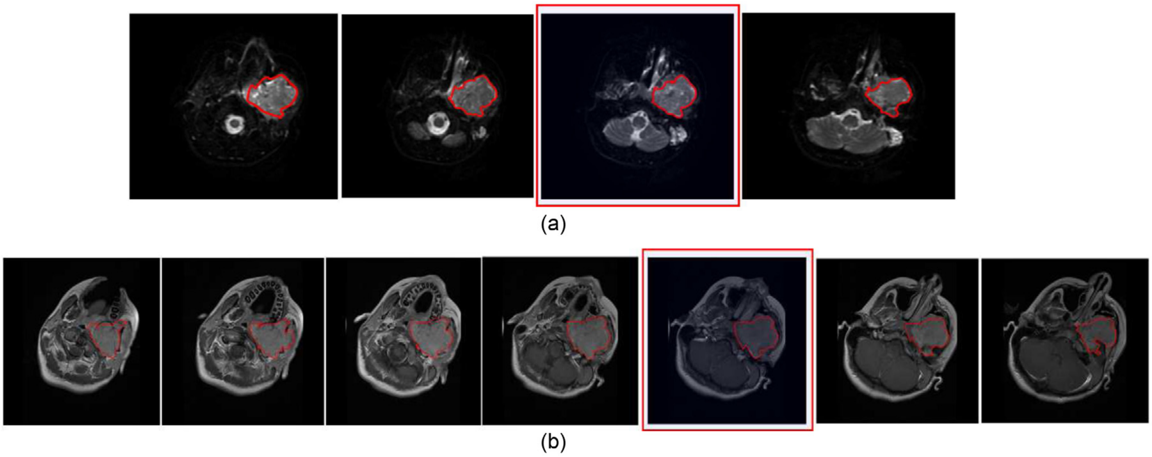

Fig. 4. Segmentation results, the slice outlined by radiologist highlighted by a red box: (a) DWI volume segmentation, (b) MRI volume segmentation. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

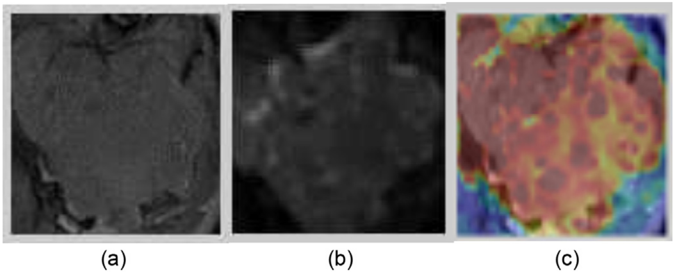

Fig. 5. Registration result: (a) tumor represented in cropped MRI; (b) tumor in cropped DWI; (c) registered image.

Using the registration and fusion process, we generate from the MRI and registered DWI volumes a stack of RGB images (227 × 227 × 7) that capture a comprehensive representation of the tumor by combining the two different MR scans.

Since the deep convolutional neural networks need to be trained on a huge number of training images to achieve satisfactory performance, we apply a standard data augmentation methodology that combines image rotation according to 3 angles and horizontal and vertical flipping techniques. After data augmentation, the dataset size increased by 12×. Fig. 6 presents augmented data for a two RGB images where the images inside the green box represents data of ERMS subtype and red box represents ARMS subtype. The rows in the figure represent flipped images and columns represent 45° and 90° rotation of each flipped image. In total, we have 1260 images of ERMS class and 504 images of ARMS class.

3.4. *Transfer learning: fine-tuning the deep convolutional neural network*

Using the augmented data (on total 1764 images), we apply the transfer learning approach (Yosinski et al., 2014) for training the popular AlexNet convolutional neural network (Krizhevsky et al., 2012) with different model training parameter values. Due to the limited data size, our goal is to fine tune the ImageNet (contains 1.2 million images with 1000 categories) pre-trained AlexNet CNN model to classify the ARMS from ERMS using the RGB fused images. Our main hypothesis is that, on a limited sized dataset, deep models that are learned via CNN transfer learning from ImageNet to other datasets of limited scales, can achieve better performance and can reduce overfitting to the small training sample despite the disparity between natural images and medical images. This is motivated by two factors. First, the earlier features of a convolutional neural network contain more generic features (e.g. edge detectors or color blob detectors) that should be useful to any image analysis tasks. This hypothesis is also promoted by the recently published articles (Sharif Razavian et al., 2014; Zhou et al., 2014). Second, it is clear that there is a huge variation between natural images and fused medical images, but the use of pre-trained weights is better than initialization of random weights since initialization of random weights has high chance of overfitting to the training data. Thus, we use transfer learning not only to replace and retrain the classifier on top of the ImageNET pre-trained AlexNet, but to also fine-tune the weights of the pretrained network by continuing the backpropagation based on the current training dataset.

The AlexNet has five convolution layers, three pooling layers, and two fully-connected layers with approximately 60 million free parameters. For RMS categorization, we change the numbers of output nodes in the last softmax classification layer. Determining the optimal learning rate for different layers for the very deep network like AlexNet is challenging. Thus, we follow the approach proposed in Zhou et al. (2014) where all CNN layers except the last two are fine-tuned at a learning rate 10 times smaller than the default learning rate. The last fully-connected layer is randomly initialized and freshly trained, in order to accommodate the new object categories (ERMS and ARMS). All the CNN layers of AlexNet except the newly modified ones are initialized with the weights of a previously trained related model and trained with a new task with a low learning rate of 0.001. The modified layers with two classes are initialized randomly, and their learning rates are set with a higher learning rate of 0.01. To “center” the data around zero mean for training, we normalize the images by subtracting mean image which computed by averaging over the training images. This typically helps the network to learn faster since gradients act uniformly.

3.5. *Model training and evaluation*

We trained the CNN on the current dataset using Matconvnet (Vedaldi and Lenc, 2015) on a 2.3 GHz Intel Core i7 with 8GB, and graphics card NVIDIA GeForce GT 650M 1024 MB. The CNN training parameters are the following: momentum 0.9, weight decay 0.0005, epoch size 50 and a batch size 50. For model training and evaluation, we applied leave-one-out cross validation (LOOCV) which is a special type of k-fold validation where we fit the model to all of the data and the number of folds equals the number of instances in the data set, i.e. 21 patients. k-Fold cross-validation is not a common approach for training the CNN since the deep CNN are usually trained on very large dataset and a portion of the data

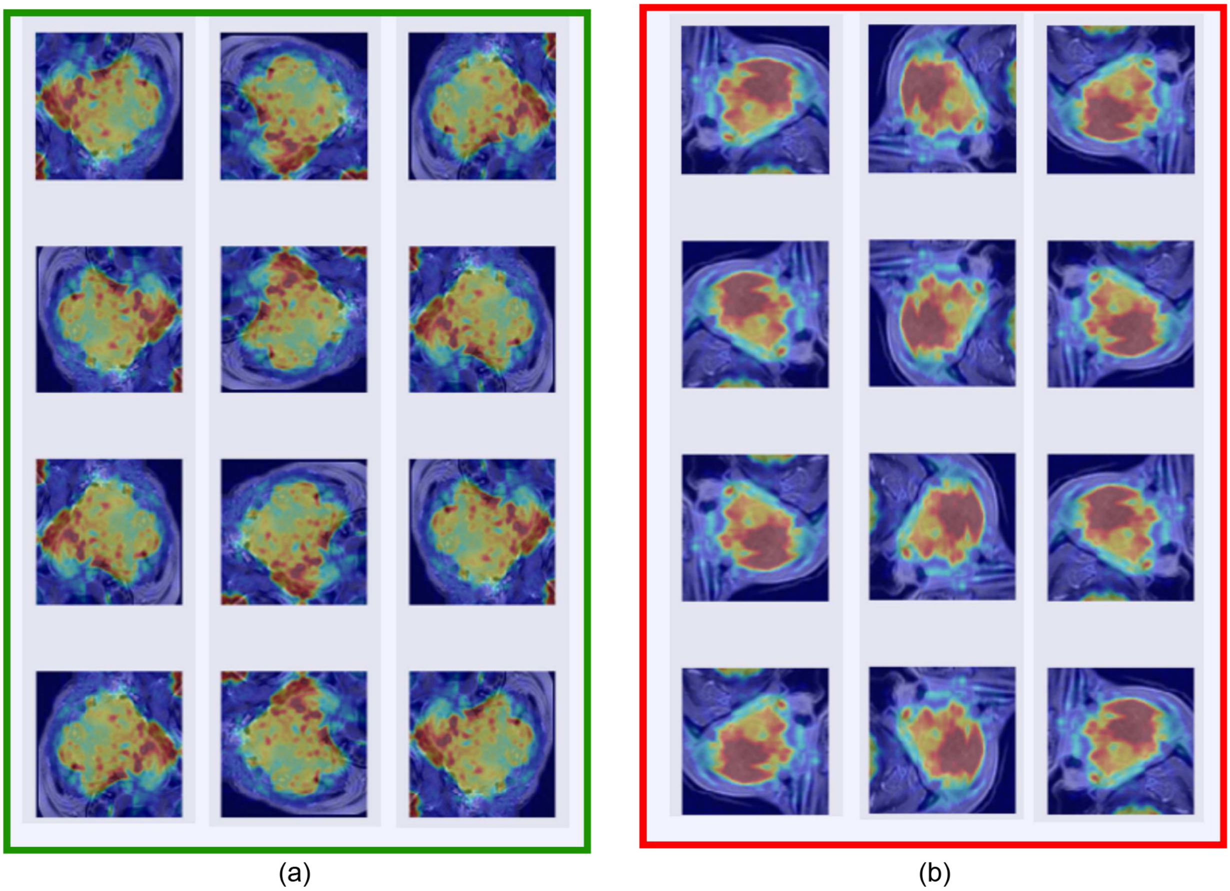

Fig. 6. Augmented data represented: (a) Embryonal rhabdomyosarcoma (ERMS) and (b) Alveolar rhabdomyosarcoma (ARMS). (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

is held out for validation. Moreover, the k-fold cross validation can be extremely expensive for training a deep CNN. However, the size of our limited dataset does not permit to hold out much data to reliably validate the model. Therefore, we used LOOCV which can give us a slightly conservative estimate of the performance of that model.

In the leave-one-out validation, we trained the CNN model where 1680 images were used as training data and 84 images were used for validation. The 2D slices of each fused volume are treated independently in the study while they may have some correlation. Thus, we created a *patient level separation* while selecting the validation set, where we train the model on the data of 20 patients and the data belongs one patient held out for validation. To evaluate our approach, we cross-validated the model for all the 21 patients. The *patient level separation* approach give us more legitimate evaluation of the model since we completely isolated the validation from the training set. We derive the patient-level classification by *majority voting* where the patient is classified to a particular class C_i if largest number of images of that patient receives classification class label C_i.

4. Results

Fig. 7 shows the performance of each epoch during training and validation of the CNN model for the last fold. The left pane shows the training error (blue) and validation error (orange) across training epochs and the right pane shows the objective function value. After 50 epochs the validation error was about 0.15 as shown in Fig. 7 and the training accuracy is about 90%. As seen from the figure, the network is saturated after 30 epochs. From leave-one-out cross validation, we achieved mean validation accuracy 85% and standard deviation ± 0.8%. The higher mean accuracy shows that the features extracted from the pre-trained AlexNet is well descriptive for the RMS classification task and the low variance is an indicator that the model is not over-fitted to the training set.

Fig. 8 shows the LOOCV results in the form of image-wise as well as patient-wise confusion matrix. Each column of the matrix represents the instances in a predicted class while each row represents the instances in an actual class. As seen from Fig. 8a, the image-wise error is evenly distributed among the two classes even if the ERMS class contains more instances compare to the ARMS. Thanks to the shuffling of the images and training with small batches, the network had opportunity to learn the discriminative features of the two classes. Fig. 8b shows the patient-wise confusion matrix derived by majority voting where only 1 patient among 21 has been predicted incorrectly. For this study, image-wise classification works as a weak classifier and combining the image-wise result through majority voting boosted the patient-wise classification performance.

In order to qualitatively understand the behavior of our CNN classifier, we visualize the weights of the first layer of the CNN which is looking directly at the RGB pixel data, and also the second layer that combines the features learnt from the first layer (see in Fig. 9). First of all, the smooth filters without any noisy patterns can be treated as an indicator of a well-trained and converged network with proper regularization strength. From the visualization of the weights, we observed that the layers learnt mainly directional edge patterns and distinct arrangements of colored blobs from the fused images. These findings closely match our expectation. Mainly because we realized from the inspection of fused images

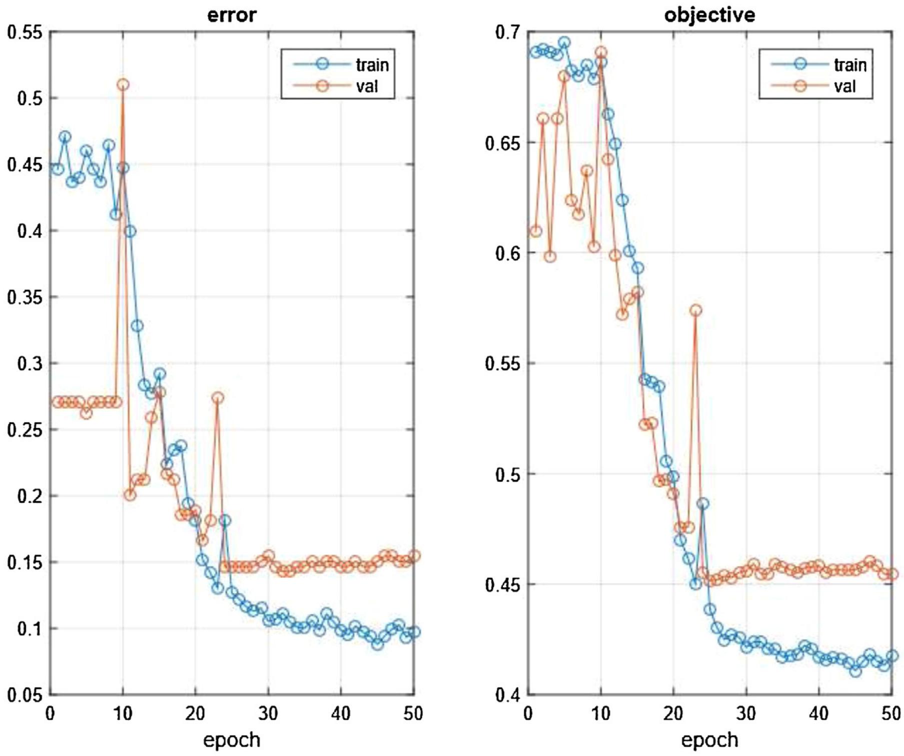

Fig. 7. Performance evaluation of epoch: error rate and objective function. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

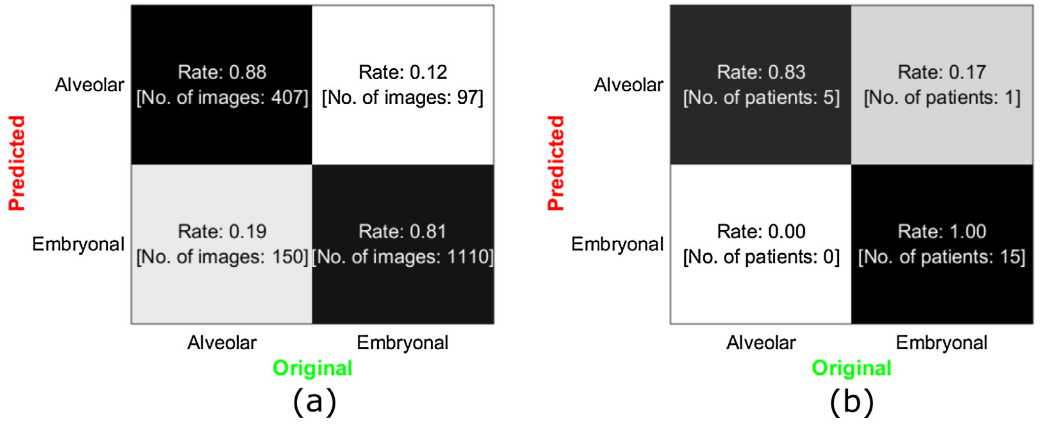

Fig. 8. Confusion matrix derived from LOOCV: (a) image-wise; (b) patient-wise, based on majority voting.

(see Fig. 6) that the patterns of colored blobs generated from diffusion weighted MR images that represent tumor cell density, and edges generated from T1 weighted MR images that represent tumor anatomy, are the most discriminative features for distinguishing ERMS and ARMS subtypes of the tumor.

5. Discussion

This paper describes a radiomics framework that distinguishes embryonal (ERMS) and alveolar (ARMS) subtypes of Rhabdomysarcoma (RMS) by analyzing two multiparametric MR images. The method takes as input contrast enhanced T1 weighted MR and diffusion weighted MR volumes, a user provided outline of the tumor on a single slice on both modalities, and produces a histopathological classification of the tumor. The proposed framework could assist the radiologists in classifying RMS subtypes directly from the multiparametric MR images and avoid the need of PET scan that requires injecting radiopharmaceutical materials in infants. The semi-automated framework is composed of multimodal image segmentation, fusion, and a transfer learning phases. Our trained model achieved 85% accuracy in classifying fused images according to the RMS subtypes. By combining image-wise classification result using majority voting, the system achieved 95% classification accuracy. The performance of the deep CNN model is appears to be at least as good as the performance of histopathologists when they evaluate biopsy specimens (Asmar et al., 1994). The study of discriminative features for classifying ERMS and ARMS subtypes may also have a beneficial impact on the diagnostic interpretation of the fused multiparametric MR images.

The main limitations of this study are that the number of cases is relatively small and this increases the potential for overfitting the model to the training dataset. We used standard data augmentation

Please cite this article in press as: Banerjee, I., et al., Transfer learning on fused multiparametric MR images for classifying histopathological subtypes of rhabdomyosarcoma. Comput Med Imaging Graph (2017), http://dx.doi.org/10.1016/j.compmedimag.2017.05.002

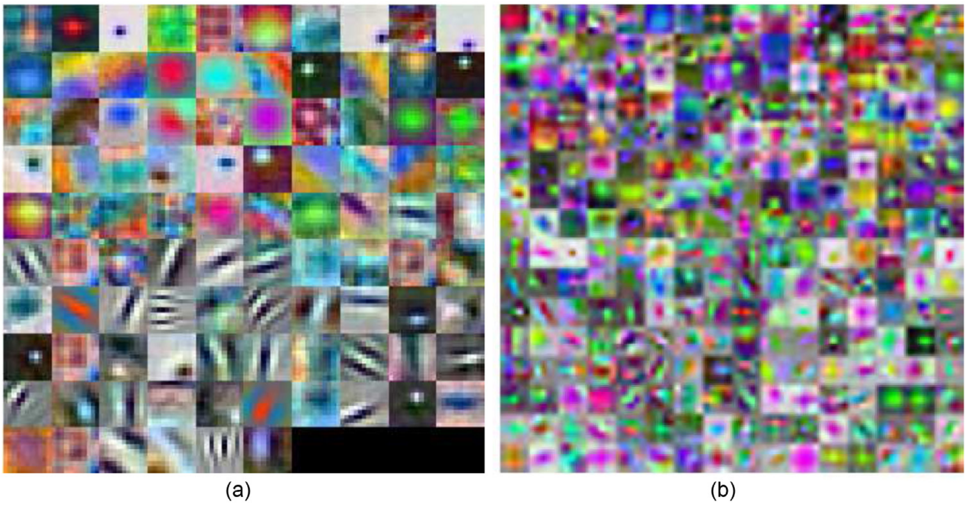

Fig. 9. Weights of the first (b) and second (b) CONV layer of the pre-trained AlexNet looking at a RGB fused images.

techniques (flip and rotation) to boost the network performance. We report the performance of our model on completely isolated test data which were selected via patient level separation (see Section 3.5). High classification accuracy achieved on the test set clearly shows that the model was not overfitted to the training set. Moreover, an imbalance in the dataset (ARMS (n = 6) and ERMS (n = 15)) could potentially penalize the accuracy of the minority class. However, in our study we have a representative sample, since ERMS comprises approximately 60–70% of childhood RMS cases while ARMS comprises approximately 20–25%. Testing the method on a larger data set will allow us to further validate the generalizability of the method.

In spite of the limited size of the current dataset, our preliminary results suggest that combination of multimodal radiological image fusion and transfer learning may be promising for RMS classification problem. Being encouraged by the results, we are planning to further investigate the RMS classification performance using a large training dataset and test various CNN architectures to evaluate the impact of the size and number of filters in the classification as well as the number of output units in the fully connected layer. Another future direction would be to test the system performance with PET scan images which may serve as a good complement of DWI scans for capturing functional activity of the tumor cells. We would also like to extend our framework to approach other clinical problem domains where the multimodal or multiparametric image fusion is expected play a significant role, such as identifying aggressive prostate cancer lesions from multimodal images.

6. Conclusion

We develop an efficient CADx system for the differential diagnosis of embryonal (ERMS) and alveolar (ARMS) subtypes of Rhabdomysarcoma (RMS). The system executes a completely automatic pipeline that performs segmentation and fusion of diffusion-weighted MR scans (DWI) and gadolinium chelate-enhanced T1−weighted MR scans (MRI) and classifies the fused images based on a trained deep CNN model. The system derives the final patient-level diagnosis by majority voting of the fused images. Required human interaction is only limited to the manual outlining of the tumor on a single slice. Our system can be exploited to provide a fast and reproducible diagnosis of RMS subtypes only by analyzing non-invasive multiparametric MR scans.

Conflict of interest statement

All of authors declare that they have no conflict of interest.

Acknowledgements

This work was supported in part by grants from the National Cancer Institute, National Institutes of Health, U01CA142555, 1U01CA190214, and 1U01CA187947.

References

Ardon, R., Cohen, L.D., 2006. Fast constrained surface extraction by minimal paths. Int. J. Comput. Vis. 69 (1), 127–136.

Asmar, L., Gehan, E.A., Newton, W.A., Webber, B.L., Marsden, H.B., Van Unnik, A.J., Hamoudi, A.B., Shimada, H., Tsokos, M., Harms, D., et al., 1994. Agreement among and within groups of pathologists in the classification of rhabdomyosarcoma and related childhood sarcomas. report of an international study of four pathology classifications. Cancer 74 (9), 2579–2588.

Bar, Y., Diamant, I., Wolf, L., Greenspan, H., 2015. Deep learning with non-medical training used for chest pathology identification. SPIE Med. Imaging Int. Soc. Opt. Photonics, 94140V.

Baum, S.H., Frühwald, M., Rahbar, K., Wessling, J., Schober, O., Weckesser, M., 2011. Contribution of PET/CT to prediction of outcome in children and young adults with rhabdomyosarcoma. J. Nucl. Med. 52 (10), 1535–1540.

Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H., et al., 2007. Greedy layer-wise training of deep networks. Adv. Neural Inf. Process. Syst. 19, 153.

Brenner, W., Conrad, E.U., Eary, J.F., 2004. FDG PET imaging for grading and prediction of outcome in chondrosarcoma patients. Eur. J. Nucl. Med. Mol. Imaging 31 (2), 189–195.

Ciompi, F., de Hoop, B., van Riel, S.J., Chung, K., Scholten, E.T., Oudkerk, M., de Jong, P.A., 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. Med. Image Anal. 26 (1), 195–202.

Freedman, D., Zhang, T., 2005. Interactive graph cut based segmentation with shape priors. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005, vol. 1, IEEE, 755–762.

Greenspan, H., van Ginneken, B., Summers, R.M., 2016. Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. IEEE Trans. Med. Imaging 35 (5), 1153–1159.

He, Y., Luo, Y., Hu, D., 2006. Semi-automatic initialization of gradient vector flow snakes. J. Electron. Imaging 15 (4), 043006.

Hoogi, A., Beaulieu, C.F., Cunha, G.M., Heba, E., Sirlin, C.B., Napel, S., Rubin, D.L., 2016. Adaptive Local Window for Level Set Segmentation of CT and MRI Liver Lesions, arXiv preprint arXiv:1606.03765.

Juntu, J., Sijbers, J., De Backer, S., Rajan, J., Van Dyck, D., 2010. Machine learning study of several classifiers trained with texture analysis features to differentiate benign from malignant soft-tissue tumors in T1-MRI images. J. Magn. Resonance Imaging 31 (3), 680–689.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst., 1097–1105.

Lahmiri, S., Boukadoum, M., 2011. Classification of brain MRI using the LH and HL wavelet transform sub-bands. 2011 IEEE International Symposium on Circuits and Systems (ISCAS), IEEE, 1025–1028.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.

Lee, H., Grosse, R., Ranganath, R., Ng, A.Y., 2011. Unsupervised learning of hierarchical representations with convolutional deep belief networks. Commun. ACM 54 (10), 95–103.

Li, C., Huang, R., Ding, Z., Gatenby, J.C., Metaxas, D.N., Gore, J.C., 2011. A level set method for image segmentation in the presence of intensity inhomogeneities with application to MRI. IEEE Trans. Image Process. 20 (7), 2007–2016.

Malempati, S., Hawkins, D.S., 2012. Rhabdomyosarcoma: review of the Children's Oncology Group (COG) soft-tissue Sarcoma committee experience and rationale for current COG studies. Pediatr. Blood Cancer 59 (1), 5–10.

Mayerhoefer, M.E., Breitenseher, M., Amann, G., Dominkus, M., 2008. Are signal intensity and homogeneity useful parameters for distinguishing between benign and malignant soft tissue masses on MR images? Objective evaluation by means of texture analysis. Magn. Resonance Imaging 26 (9), 1316–1322.

Ognjanovic, S., Linabery, A.M., Charbonneau, B., Ross, J.A., 2009. Trends in childhood rhabdomyosarcoma incidence and survival in the United States, 1975–2005. Cancer 115 (18), 4218–4226.

Othman, M.F., Basri, M.A.M., 2011. Probabilistic neural network for brain tumor classification. 2011 Second International Conference on Intelligent Systems, Modelling and Simulation (ISMS), IEEE, 136–138.

Othman, M.F.B., Abdullah, N.B., Kamal, N.F.B., 2011. MRI brain classification using support vector machine. 2011 4th International Conference on Modeling, Simulation and Applied Optimization (ICMSAO), IEEE, 1–4.

Porter, T., Duff, T., 1984. Compositing digital images. ACM SIGGRAPH Computer Graphics, vol. 18, ACM, 253–259.

Rubin, D.L., Willrett, D., O'Connor, M.J., Hage, C., Kurtz, C., Moreira, D.A., 2014. Automated tracking of quantitative assessments of tumor burden in clinical trials. Transl. Oncol. 7 (1), 23–35.

Sharif Razavian, A., Azizpour, H., Sullivan, J., Carlsson, S., 2014. CNN features off-the-shelf: an astounding baseline for recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 806–813.

van Ginneken, B., Setio, A.A., Jacobs, C., Ciompi, F., 2015. Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans. 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), IEEE, 286–289.

Vedaldi, A., Lenc, K., 2015. Matconvnet: convolutional neural networks for matlab. Proceedings of the 23rd ACM International Conference on Multimedia, ACM, 689–692.

Vercauteren, T., Pennec, X., Perchant, A., Ayache, N., 2009. Diffeomorphic demons: Efficient non-parametric image registration. NeuroImage 45 (1), S61–S72.

Vese, L.A., Chan, T.F., 2002. A multiphase level set framework for image segmentation using the Mumford and Shah model. Int. J. Comput. Vis. 50 (3), 271–293.

Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How transferable are features in deep neural networks? Adv. Neural Inf. Process. Syst., 3320–3328.

Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., Oliva, A., 2014. Learning deep features for scene recognition using places database. Adv. Neural Inf. Process. Syst., 487–495.