2015

A Comparative Study of Two Prediction Models for Brain Tumor Progression

Deqi Zhou
Princess Anne High School

Loc Tran
Old Dominion University

Jihong Wang
University of Texas MD Anderson Cancer Center

Jiang Li
Old Dominion University, jli@odu.edu

Karen O. Egiazarian (Ed.)

See next page for additional authors

Original Publication Citation

Zhou, D., Tran, L., Wang, J., & Li, J. (2015). A comparative study of two prediction models for brain tumor progression. In K. O. Egiazarian, S. S. Agaian, & A. P. Gotchev (Eds.), *Image Processing: Algorithms and Systems XIII, Proceedings of SPIE-IS&T Electronic Imaging, SPIE Vol. 9399* (93990W). SPIE. https://doi.org/10.1117/12.2082645

A Comparative Study of Two Prediction Models for Brain Tumor Progression

Deqi Zhou[1], Loc Tran[2], Jihong Wang[4] and *Jiang Li[2,3]
[1]The IB Program Princess Anne High School 4400 Virginia Beach Boulevard, Virginia Beach, VA 23462, E-mail: deqizhou@gmail.com
[2]Department of Electric and Computer Engineering, Old Dominion University, Norfolk, VA 23508, E-mail: JLi@odu.edu
[3]Guangxi Key Laboratory for Spatial Information and Geomatics
Guilin University of Technology, Guilin, China
[4]Department of Imaging Physics, the University of Texas MD Anderson Cancer Center, Houston, TX 77030, E-mail: Jihong.Wang@mdanderson.org

ABSTRACT

MR diffusion tensor imaging (DTI) technique together with traditional T1 or T2 weighted MRI scans supplies rich information sources for brain cancer diagnoses. These images form large-scale, high-dimensional data sets. Due to the fact that significant correlations exist among these images, we assume low-dimensional geometry data structures (manifolds) are embedded in the high-dimensional space. Those manifolds might be hidden from radiologists because it is challenging for human experts to interpret high-dimensional data. Identification of the manifold is a critical step for successfully analyzing multimodal MR images.

We have developed various manifold learning algorithms (Tran et al. 2011; Tran et al. 2013) for medical image analysis. This paper presents a comparative study of an incremental manifold learning scheme (Tran. et al. 2013) versus the deep learning model (Hinton et al. 2006) in the application of brain tumor progression prediction. The incremental manifold learning is a variant of manifold learning algorithm to handle large-scale datasets in which a representative subset of original data is sampled first to construct a manifold skeleton and remaining data points are then inserted into the skeleton by following their local geometry. The incremental manifold learning algorithm aims at mitigating the computational burden associated with traditional manifold learning methods for large-scale datasets. Deep learning is a recently developed multilayer perceptron model that has achieved start-of-the-art performances in many applications. A recent technique named "Dropout" can further boost the deep model by preventing weight coadaptation to avoid over-fitting (Hinton et al. 2012).

We applied the two models on multiple MRI scans from four brain tumor patients to predict tumor progression and compared the performances of the two models in terms of average prediction accuracy, sensitivity, specificity and precision. The quantitative performance metrics were calculated as average over the four patients. Experimental results show that both the manifold learning and deep neural network models produced better results compared to using raw data and principle component analysis (PCA), and the deep learning model is a better method than manifold learning on this data set. The averaged sensitivity and specificity by deep learning are comparable with these by the manifold learning approach while its precision is considerably higher. This means that the predicted abnormal points by deep learning are more likely to correspond to the actual progression region.

Keywords: Magnetic resonance imaging, Manifold learning, Deep learning, Brain tumor diagnosis

1. INTRODUCTION

The assessment and prediction of cancer progress are conventionally divided into three phases (Brunetti et al. 1996): (a) detecting and characterizing the lesion location; (b) assessing local extension of the disease and presence of metastasis and (c) tracing the tumor progress and changes over time, while responding to treatment. Both MRI and computed tomography (CT) images have been widely applied in detection and assessment of the local extension of the primary intracranial tumors and are routinely used for diagnosis of secondary tumors during staging procedures in patients with primary body tumors.

Recent advancements in imaging technology result in much easier acquisition of multimodal, large-scale and heterogeneous medial datasets. For example, the current brain MR imaging incorporate MR DTI on a frequent basis for brain tumor diagnosis and treatment monitoring. Along with traditional T1, T2, or FLAIR weighted MRI scans, these high-dimensional data provide rich information and potential for high qualitative brain tumor diagnosis and treatment management.

Corresponding author: Jiang Li jli@odu.edu

Image Processing: Algorithms and Systems XIII, edited by Karen O. Egiazarian, Sos S. Agaian, Atanas P. Gotchev, Proc. of SPIE-IS&T Vol. 9399, 93990W · © 2015 SPIE-IS&T · CCC code: 0277-786X/15/$18 · doi: 10.1117/12.2082645

However, interpreting these large-scale and high-dimensional datasets is challenging (Araki et al. 1984). For example, to examine malignant brain tumors, ADC values, quantitative maps of apparent diffusion coefficient derived from DTI imaging, have been reported as useful indicators in distinguishing of tumor tissue from the surrounding edema by Sinha et al (Sinha, Bastin et al. 2002). Other researchers found that the differentiation has not always been successful (Kono, Inoue et al. 2001). Lam et. al reported that ADC values were not useful in identifying tumor types (Lam, Poon et al. 2002). These claims have been challenged by a number of researchers where results showed that high-grade gliomas have lower ADC values while low-grade ones are opposite (Kono, Inoue et al. 2001; Sadeghi, Camby et al. 2003). Analysis performed on diffusion anisotropic (FA) values also showed contradictory results in its ability for differentiation of enhancing tumor from edematous brain cancer (Bastin et al. 2002). Similar situations can also be found in meningiomas diagnosis. A significant difference in peritumoral ADC and FA values between low-grade meningiomas and high-grade gliomas has been reported by Bastin et al (Bastin, Sinha et al. 2002), possibly reflecting the presence of tumor-infiltrated edema in gliomas. On the other hand, Lu et al. showed that there is no statistically significant difference of ADC and FA values between intra- and extra-axle lesions (Lu, Ahn et al. 2004). No significant difference was found for ADC values between peritumoral hyperintense regions or peritumoral normal-appearing white matter and high-grade gliomas (Provenzale, McGraw et al. 2004).

We suspect that one or few MRI scans may not have enough discriminating power to reliably differentiate various tissues. We have developed a large-scale manifold learning strategy to interpret the highly correlated multiple MRI scans for brain tumor diagnosis (Tran et al. 2011; Tran et. al., 2013). Our results indicate that though an individual MRI scan is probably insufficient to reliably distinguish different tissue types, taking advantage of multiple MRI images may lead to a successful classification. The assumption is that though these MRI scans or derived volumes such as ADC have different imaging parameters, these scans are highly correlated and the number of independent variables embedded in the data is less than the number of scans. Effective identification of the independent variables is critical for tumor diagnosis because the side effect from unrelated information can be mitigated. These independent variables are usually called manifolds in machine learning community and are identified by manifold learning. One challenge of manifold learning is that there is a matrix Eigen value decomposition step and the size of the matrix is n by n, where n is the number of data points, which is not feasible for large data sets. We have developed an incremental manifold learning system for brain tumor progression prediction to overcome this challenge (Tran et. al., 2013).

Deep learning (Hinton et al, 2006) is a new algorithm in machine learning community. The primary purpose of deep learning is to automatically extract features for various applications including computer vision, speech recognition and many other pattern recognition tasks. In this paper, we will compare our incremental manifold learning algorithm with deep learning in the application for brain tumor diagnosis and progression prediction. We have collected patient data at the MD Anderson Cancer Center and retrospectively studied four brain tumor patients. The purpose of this study is to evaluate if deep learning can provide advantages over traditional manifold learning.

2. OVERVIEW OF DEEP LEARNING AND MANIFOLD LEARNING MODELS

In this section, we will review the two learning models including deep learning and incremental manifold learning. The incremental manifold learning model was developed by our group and we utilized the deep learning code developed by Hinton's group in this study.

2.1 Deep learning

Deep learning is a method to effectively train a Multi-Layer Perceptron (MLP) with more than one hidden layer (deep structure). Compared to MLP with one hidden layer, the deep structure can approximate a complex nonlinear input/output mapping with fewer parameters. However, when training a deep structure with the traditional back propagation (BP) method, the training process often gets stuck in local minima if the network is randomly initialized. On the other hand, deep learning has a mechanism of layer-by-layer unsupervised training to provide an improved parameter initialization for the BP algorithm. It has been experimentally confirmed that the deep learning process often reaches better local minima (Dumitru et al. 2010; Bengio et al. 2013).

Deep learning is motivated by intuition, circuit theory and neuroscience (Dumitru et al. 2010). One example is from circuit theory. A "shallow" circuit may represent the same function, but it will be exponentially wider than a "deep" circuit (lots of duplications of logical gates). Similarly, a shallow machine learning architecture would involve a lot of duplications of effort to express the learned objects and such architecture has been shown to suffer

from the problem of over-fitting, which leads to a poor generalization capability. Instead, deep architecture could more gracefully reuse previous computations and discover complicated relations of the input with a deep model structure. The deep learning framework is also mimicking what a human brain does for classification and recognition. In human vision system, information is processed layer by layer. Each layer achieves different level of abstraction with the first layer detects low level characteristics such as edges in images and the highest layer extract abstract objects like faces by integrating lower level representations from previous layers. Finally, our brain tries to identify them by comparing the integrated objects with those we saw before.

With the motivation to utilize a deep architecture, to train a deep architecture, the standard way is to use a supervised learning algorithm. However, it is shown that feedback error messages from output layer become progressively noisier as it is passed to lower levels, and thus are less and less effective for weights adaptation (Dumitru et al. 2010). The difficult to train a deep architecture has been the major hurdle to the application of deep neural network structure in the machine learning community. In 2006, a breakthrough in deep learning has made the deep architecture training possible. To improve the signal-to-noise ratio at lower levels, there are two main features need to be performed: unsupervised learning and the creation of features one layer at a time (Hinton et al. 2006; Yoshua et. al 2009, Dumitru et al. 2010). With unlabeled data, the deep structure learns multiple layers of nonlinear representations for data in an unsupervised pre-training step and achieves a good initialization of parameters in a layer-by-layer manner (Fig. 1 left). The model parameters are then fine-tuned with labels for prediction or classification (Fig. 1 right, red dots on top). Enabled by these two features, deep learning can achieve better performance (better local minima) with a simple yet deep architecture.

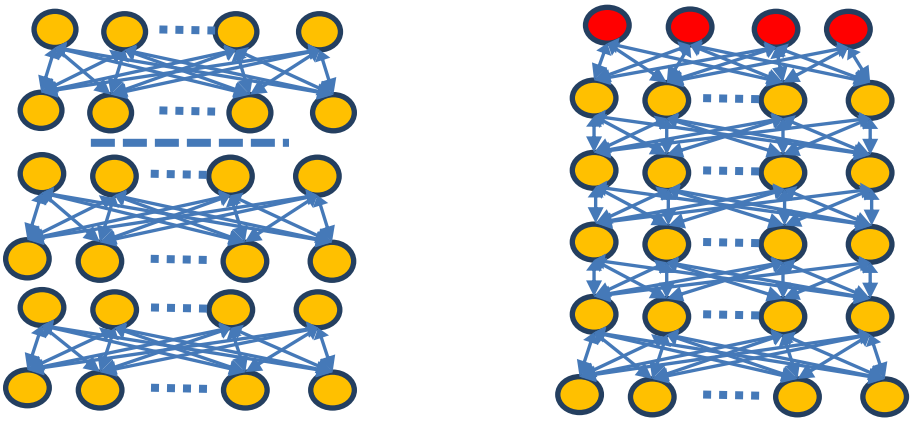

Figure 1. Deep learning framework. Left: Layer by layer unsupervised pre-training.
Right: weight fine-tune with labels shown as red.

2.1.1 Unsupervised Learning

With unsupervised learning, deep learning is trying to understand the data first, i.e., to obtain a task specific representation for the data so that a better classification can be achieved, which is similar to recognition in human brain. If the data label is provided, the deep learning framework has the ability to refine those representations making them task specific. It has experimentally proven that the unsupervised learning step in deep learning plays a critical role in the success of the training of deep learning (Dumitru et al. 2010).

2.1.2 Creation of features one layer at a time

In each lower level, the objective is to train a model such that outputs (features) can keep as much information as input. In this way, the features can be more and more abstract, while keeping the original sensor information as much as possible. At the highest level, the limited label information (known outputs) can be used to fine tune the parameters and further pass the feedback through the architecture layer by layer. With these two tricks, deep learning is able to train a deep architecture relatively easily and achieve a good generalization capability. It is also worth noting that the deep learning architecture can simultaneous model multiple outputs. When learning multiple tasks simultaneously, deep learning will achieve a set of shared data representations for the tasks that better handle correlations among them leading to a nature multi-task learning framework.

2.2 Incremental Manifold Learning

The diagram of the incremental manifold learning system is shown in Fig. 2, left. We first select a subset of the datasets as landmarks based on statistical sampling methods. A skeleton of the manifold embedded in the high-dimensional datasets is then identified from the landmarks, and out-of-bag samples are inserted into the skeleton (Tran et al. 2011; Tran et al. 2013). We briefly describe each component in the system as follows.

2.2.1. Sampling based on Local Tangent Space Variation (LTV)

To keep a faithful representation of the original manifold, landmarks should be carefully selected from the original data. Ideally, landmarks should be the smallest subset that can preserve the geometry in the original data. Fig. 2 right shows a toy dataset to illustrate the basic idea of LTV. Heuristically, to preserve the data structure after sampling, we should keep more data points near area 'A' rather than area 'B' in Fig. 2 right because data structures near 'A' change more abruptly. Based on this observation, we assign an importance value for each of the points based on its local tangent space variation, i.e., how quick the directions of the red arrows in Fig. 2 right change along the data manifold. The local tangent space variation can be computed by performing a PCA analysis in the local neighborhood of each data point (Tran, et al. 2013). The normalized importance values are then used to guide the subsequent statistical sampling for landmark selection.

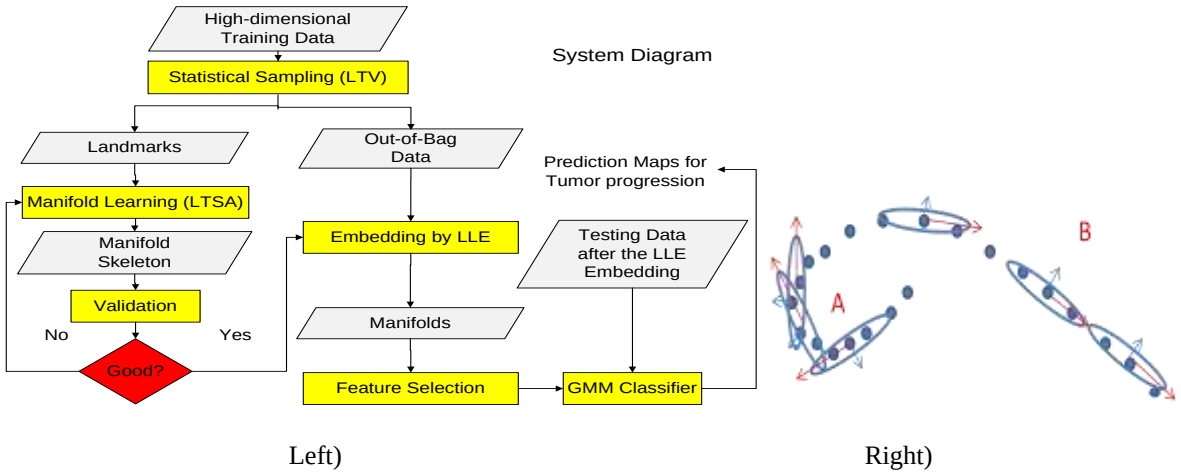

Figure 2. Left: System diagram of the proposed method and Right: a detailed illustration of local linear embedding for the remaining data points, where the yellow dot represents one remaining data sample to be embedded.

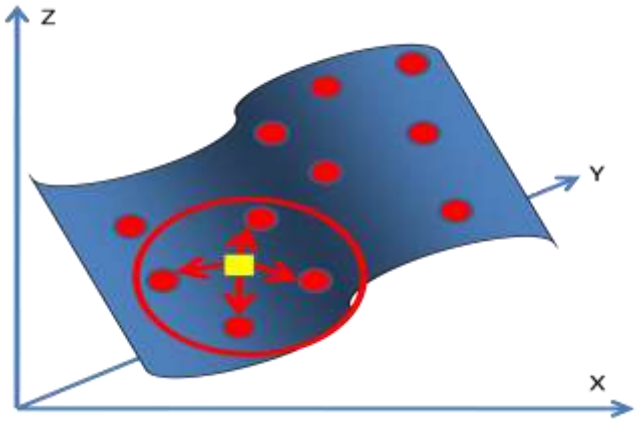

Figure 3. Local Linear Embedding.

2.2.2. Embedding by LLE

Once the manifold skeleton is learned, we utilize the LLE algorithm (Roweis and Saul 2000; Donoho and Grimes 2005) to insert the remaining data points into the manifold skeleton as shown in Fig. 3), where the red dots are landmarks consisting of the manifold skeleton, and the yellow square is a remaining data point to be embedded into the skeleton. There are three steps involved in this task. 1) Discover K nearest landmarks in the original data space for the yellow square, 2) Compute a linear model that best reconstructs the data point using the K landmarks and 3) insert the yellow square into the skeleton by reusing the reconstruction weights in the linear model (Tran et al. 2011; Tran et al. 2013).

2.3. Feature selection and classifier training

The learned low-dimensions are usually not equally important for the subsequent tumor growth prediction. Feature selection techniques such as the simple Fisher score was utilized to select relevant dimensions for the subsequent prediction. In the experiments, the Fisher scores for features beyond the third highest were found to be two orders of magnitude smaller than the highest score and were then discarded. Finally, the data set was applied to a Gaussian mixture model (GMM) for prediction/classification. A benefit of using GMM is that a posterior probability mapping can be generated. The training samples used were the same points that were selected in the LCV sampling step. And the trained GMM model was applied to all remaining points, a probability map can be produced for the MRI data set. The final classification can be computed by thresholding the GMM classification probability.

3. EXPERIMENTS AND ANALYSIS

3.1 Data Preparation

MRI scans including FLAIR, T1-weighted, post-contrast T1-weighted, and DTI were collected from four patients with progressing brain tumors at MD Anderson Cancer Center. Five scalar volumes which include apparent diffusion coefficient (ADC), fractional anisotropy (FA), max eigenvalues, middle eigenvalues, and min eigenvalues were also computed from the DTI volume yielding a total of ten image volumes for each visit per patient. Each patient went through a series of clinical visits over the course of two years. Using the vtkCISG toolkit, a strict form of registration was applied to each patient that aligned all volumes to the DTA volume (Tran et al. 2011; Tran et al. 2013). After registration, each pixel location can be represented by a ten-dimensional feature vector corresponding to the ten MRI scans. One visit was selected and labeled as "Visit 1," and a later visit that showed evidence of tumor progression was selected and labeled "Visit 2." A radiologist defined the tumor regions on the post-contrast T1-weighted and FLAIR scans, respectively. For training purposes, normal regions far from the tumor regions have also been defined. Figure 4 shows the selected regions for one subject for both visits. The yellow dotted region is the normal region while the red polygon is the abnormal region. Note that the abnormal region is larger at Visit 2. The goal of this study is to predict the progressed region at Visit 2.

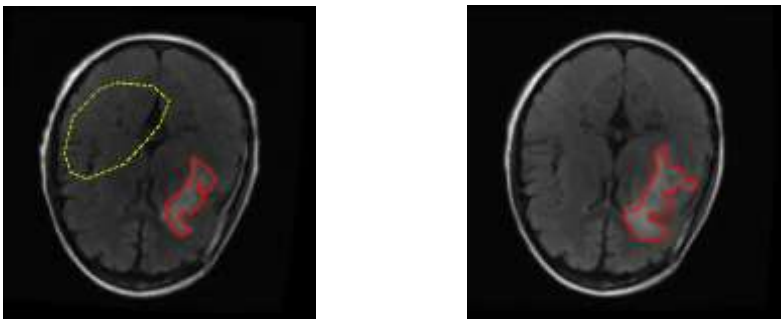

Figure 4**.** Sample tumor and normal regions where the red tumor regions are labeled by a radiologist and the yellow region denotes a normal region. FLAIR images at visit 1 (left) showing a progressed tumor at visit 2 (right).

Figure 5. Cropped images of sample abnormal classification regions by the deep neural network. The red outlines denote radiologist marked tumor regions at Visit 2.

3.2 Experiments and Results

Both models were trained on a sampled set from both the abnormal and normal region at Visit 1. The output of the models will be the classification of normal and abnormal regions that will be mapped to Visit 1 and Visit 2 to compute sensitivity, specificity and precision at the two visits. There are three hidden layers in the deep structure having a configuration of [500 100 20] for hidden units in each layer. For manifold learning, a Gaussian mixture model (GMM) classifier is trained using the first three dimensions ranked by the Fisher score.

Table 1 show quantitative performance metrics calculated as an average over 4 subjects. The sensitivity measures the ratio between the numbers of pixels correctly predicted as abnormal versus the total number of marked abnormal pixels. This measure was calculated for both Visit 1 and Visit 2. Specificity is the ratio of the correctly predicted normal tissue samples inside the normal contours. The precision is the number of correctly predicted abnormal pixels divided by the total number of predicted abnormal points. The precision will be high if every pixel predicted as abnormal is within the marked abnormal region and conversely, the metric will be low for methods that have an over-estimated tumor region. The precision was calculated only at Visit 2 because the abnormal region was expected to expand between Visit 1 and Visit 2. The results for Raw are found by directly applying a GMM classifier in the high dimensional space. For PCA, the dimensionality reduction is performed using principal component analysis and the dimension was reduced to three.

Figure 5 shows some sample classification regions by the deep neural network. The red outline is the marked abnormal regions at Visit 2. By comparing each method, it can be seen that the deep neural network method is an overall better method than the other techniques for this data set. The sensitivity and specificity are comparable with the manifold learning approach while the precision is considerably higher. This means that the predicted abnormal points are more likely to correspond to the actual progression region.

Table 1**.** Sensitivity, specificity and precision comparison among different methods

	Sensitivity at Visit 1	Specificity at Visit 1	Sensitivity at Visit 2	Precision at Visit 2	Average
Deep Neural Network	0.948	1.000	0.653	0.858	**0.864**
Manifold Learning	0.951	1.000	0.663	0.781	0.849
PCA	0.945	1.000	0.649	0.473	0.766
RAW	0.917	0.872	0.705	0.617	0.778

4. CONCLUSION

It is important to establish useful quantitative parameters to accurately diagnose tumor growth and radiation injury in patients who have undergone radiotherapy and subsequently manifested increased enhancement. Our preliminary results showed that both the manifold learning and deep neural network models produced better results compared to using raw data and PCA. We also showed that a more robust tumor progression model can be achieved based on deep learning compared to manifold learning. The recent technique named "Dropout" incorporated into deep learning illustrated super performances for many benchmark data sets (Hinton et al. 2012). We are investigating if the "Dropout" technique can also further improve the performance on our date set.

REFERENCES

Araki, T.; Inouye, T.; Suzuki, H.; Machida, T.; Iio, M. (1984). Magnetic resonance imaging of brain tumors: measurement of T1. Work in progress, Radiology, vol. 150, no. 1, pp. 95-8.

Bastin, M. E., S. Sinha, et al. (2002). "Measurements of Water Diffusion and T1 Values in Peritumoural Oedematous Brain." Neuroreport 13: 1335-40.

Bengio, Y., Courville, A., Vincent, P., Representation learning: A review and new perspectives, PAMI, 35(8), 1798-1828 (2013).

Brunetti, Arturo, Bruno Alfuno, Andrea Soticelli, Enrico Tedeschi, Ciro Mainolfi, Eugenio M. Covelli, Luigi Aloj, Maria Rosaria Punico, Lucia Bazzicalupo and Marco Salwatore (1996). Functional characterization of brain tumors: an overview of the potential clinical value, Nuclear Medicine & Biology, vol. 23, pp. 699-715.

Dumitru Erhan, Yoshua Bengio, Aaron Courville, Pierre-Antoine Manzagol, Pascal Vincent, Samy Bengio, "Why Does Unsupervised Pre-training Help Deep Learning?", Journal of Machine Learning Research, vol. 11, pp. 625-660, 2010.

Donoho, D. L. and C. Grimes (2005). "Hessian Eigenmaps: New Locally Linear Embedding Techniques for High-dimensional Data." Proc. National Academy of Sciences (PNAS) 102(21): 7426-7431.

Hinton, G. E. and R. R. Salakhutdinov, Reducing the dimensionality of data with neural networks" Science, 313(5786): 504-507, 2006.

Hinton, G.E., Srivastave, N., Krizhevsky, A.,Sutskever, I., and Salakhutdinov, R. R. Improving neural networks by preventing co-adaptation of feature detectors. arXiv:1207.0580 (2012).

Kono, K., Y. Inoue, et al. (2001). "The Role of Diffusion-weighted Imaging in Patients with Brain Tumors." Am J. Neuroradiol 22: 1081-8.

Lam, W. W., W. S. Poon, et al. (2002). "Diffusion MR Imaging in Glioma: Does it Have Any Role in the Preoperation Determination of Grading of Glioma?" Clin. Radiol. 57: 219-25.

Tran, Loc; Banerjee, Deb; Sun, Xiaoyan; Wang, Jihong; Kumar, Ashok J.; Vinning, David; McKenzie, Frederic D.; Li, Yaohang; Li, Jiang (2011). A large-scale manifold learning approach for brain tumor progression prediction, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7009 LNCS, pp. 265-272.

Tran, L., Banerjee, D., Wang, J., Li, J. et. al., "High-dimensional MRI data analysis using a large-scale manifold learning approach", Machine Vision and Applications, vol. 24, no. 5, pp. 995-1014, 2013.

Lu, S., D. Ahn, et al. (2004). "Diffusion-tensor MR Imaging of Intracranial Neoplasia and Associated Peritumoral Ecdema: Introduction of the Tumor Infiltration Index." Radiology 232: 221-8. Magazine 2(6): 559-572.

Provenzale, J. M., P. McGraw, et al. (2004). "Peritumoral Brain Regions in Gliomas and Meningiomas: Investigation with Isotropic Diffusion-weighted MR Imaging and Diffusion-tensor MR Imaging." Radiology 232: 451-60. Roweis, S. T. and L. K. Saul (2000). "Nonlinear dimensionality reduction by locally linear embedding." Science 290(5500): 2323-6.

Roweis, S. T. and L. K. Saul (2000). "Nonlinear dimensionality reduction by locally linear embedding." Science 290(5500): 2323-6.

Sadeghi, N., I. Camby, et al. (2003). "Effect of Hydrophilic Components of the Extracellular Matrix on Quantifiable Diffusion-weighted Imaging of Human Gliomas: Preliminary Results of Correlationg Apparent Diffusion Coefficient Values and Hyaluronan Expression Level." Am J. Roentgenol. 181: 235-41.

Sinha, S., M. E. Bastin, et al. (2002). "Diffusion Tensor MR Imaging of High-grade Cerebral Gliomas." Am J. Neuroradiol 23: 520-7.

Yoshua Bengio "Deep Learning for Artificial Intelligence," Foundations and Trends(R) in Machine Learning, ISBN-10: 1601982941, 2009.