DESICCATION DIAGNOSIS IN LUMBAR DISCS FROM CLINICAL MRI WITH A PROBABILISTIC MODEL

*Raja' S. Alomari, Jason J. Corso, Vipin Chaudhary**

Computer Science and Engineering Dept.
University at Buffalo
Buffalo, NY 14260

Gurmeet Dhillon, MD

Proscan of Buffalo
Buffalo, NY 14221

ABSTRACT

Lumbar Intervertebral disc diseases are the main cause of lower backpain (LBP). Desiccation is a common disease resulting from different causes and ultimately most people are affected by desiccation at some age. We automate the diagnosis of desiccation by processing localized lumbar intervertebral discs. We utilize a Gibbs distribution to model the appearance and context of intensity for the desiccated discs. We use 55 clinical T2-weighted MRI for lumbar area and achieve over 96% accuracy on a cross validation experiment.

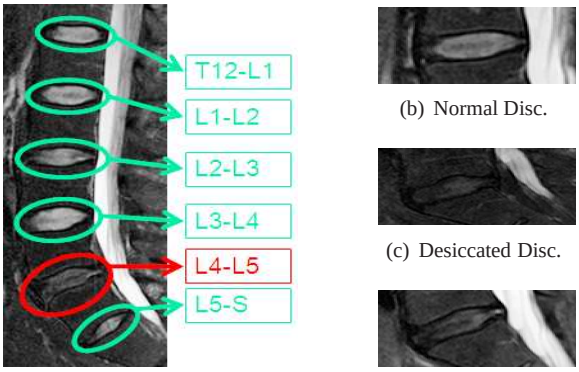
Index Terms— Computer Aided Diagnosis, MRI, Lumbar discs, Desiccation Diagnosis.

1. INTRODUCTION

Full or partial automation of diagnosis for lumbar degenerative disc diseases, *e.g.* disc desiccation in this paper, is a useful step in helping the radiologist achieve his task given the high demand on diagnosis for lower backpain (LBP). According to the National Institute of Neurological Disorders and Stroke (NINDS), LBP is the second most common neurological ailment in the United States after the headache [1]. Americans spend at least $50 billion each year on lower backpain (LBP) and over 12 million Americans have some sort of Intervertebral Disc Disease (IDD) [1].

Magnetic Resonance Imaging (MRI) are the only acceptable modalities for diagnosis of disc degeneration though radiographic data are sometimes acceptable [2]. Disc signal intensity in T2-weighted MRI is the most sensitive sign for intervertebral disc degeneration [2, 3] and especially diagnosis of disc desiccation [3]. However, due to the magnetic field inhomogeneities, disc signals (intensities) vary for reasons other than the difference in water contents [4]. Furthermore, this signal is also affected by the MRI protocol. This led radiologists to measure the disc signal with respect to an adjacent intra-body reference [4, 5]. Cerebrospinal fluid (CSF) is usually the standard reference for this purpose in the lumbar spine [3, 6, 7]. Recently, the spine signal is suggested for this purpose as well [8].

*Send correspondence to ralomari@buffalo.edu. This work was supported in part by the New York State Foundation for Science, Technology and Innovation (NYSTAR).



(a) Labeled Discs for a T2-Weighted MR for Lumbar area. Level $\mathcal{L}4 - \mathcal{L}5$ is a desiccated disc.

(b) Normal Disc.

(c) Desiccated Disc.

(d) Desiccated Disc.

Fig. 1. A Sample desiccated case and zooms for normal and desiccated discs. The relative intensity of the signal between the spine and the disc is the main criteria for diagnosis of desiccation. Images are enhanced for better visual quality.

Intervertebral disc disease diagnosis from MRI requires labeling of discs and then detection of the abnormality. In our previous work [9], we have developed a probabilistic model for localization of the six discs in sagittal T2-weighted MRI for the lumbar area. In our model, we incorporate two levels of information: low- and high-level. In the low-level, we model the local pixel properties of discs, such as appearance. In the high-level, we capture the object-level geometrical and contextual relationships between discs. We estimate the model parameters from manually labeled cases (supervised learning). We tested our model using a dataset of 20 normal cases and showed the extension to an abnormal case. However, in this paper, we use a dataset of 55 clinical cases that contains desiccated as well as normal discs. This dataset has variabilities in patient ages (17 to 81 years old), and patient heights which affect the size and appearance of the discs. Fig. 1(a) shows a sample sagittal view for a case with desiccated discs.

In this paper, we propose a method for detection of desiccated discs n_i^* in the lumbar area at each disc level i (Fig. 1(a)):

$$n_i^* = \arg\max_{n_i} P(n_i|d_i, \sigma_{\mathtt{I}(d_i)}) \quad (1)$$

where n_i is a binary random variable stating whether it is a desiccated or a normal disc and $n_i \in \mathcal{N} = \{n_i : 1 \leq i \leq 6\}$, $d_i \in \mathcal{D} = \{d_i : 1 \leq i \leq 6\}$ is the location of each lumbar disc, and $\sigma_{\mathtt{I}(d_i)}$ is the intensity of a pixel neighborhood surrounding the disc level (i).

The remainder of this paper is organized as follows. The background and related work is discussed in section 2. Then we discuss our proposed model in section 3. We then discuss our experimental settings and results in section 4. Section 5 concludes this paper.

2. BACKGROUND

Intervertebral discs are unique structures that absorb shocks between adjacent vertebrae. They connect the vertebrae and act as the pivot point which allows the spine mobility by bending and rotating. They make about one fourth of the spinal column length [10]. A disc is composed of two parts: an outer strong ring called annulus fibrosis surrounding a soft gel-like inner called nucleus pulposus. The nucleus pulposus consists of 80% to 85% water in normal cases.

Disc desiccation is a degenerative disc disease that dries up the water contents. It leads to breakage and arthritis. If it progresses, it may bulge and possibly cause pressure onto the spinal nerves. Most people at some age are diagnosed with desiccation because of normal aging. However, there are many other reasons for disc desiccation such as trauma, accidents, suddent weight loss, and many others that are unknown [10].

Computer aided diagnosis (CAD) in intervertebral disc diseases (IDD) has been attracting many researchers over the last three decades [11, 12, 13]. In the mid 1980s, Jenkins et al. [14] performed a valuable analysis study on 107 normal and 18 abnormal cases. They analyzed the relation between proton density and age in normal discs. They concluded that quantitative MR analysis may assist in the diagnosis of intervertebral disc degeneration.

Bounds et al., [11] utilized a neural network for diagnosis of backpain and sciatica. They have three groups of doctors perform diagnosis as their validation mechanism. They claimed that they achieve better accuracy than the doctors in the diagnosis. However, the lack of data forbade them from full validation of their system. Similarly, Vaughn [12] conducted a research study on using neural network (NN) for assisting orthopaedic surgeons in the diagnosis of lower back pain. Lower backpain is classified into three broad clinical categories: Simple Low Back Pain (SLBP), Root Pain (ROOTP), and Abnormal Illness Behaviour (AIB). About 200 cases were collected over the period of 2 years with diagnosis from radiologists. Twenty five features were used to train the NN including symptoms clinical assesment results. The NN achieved 99% of training accuracy and 78.5% of testing accuracy.

Tsai et al., [13] used geometrical features (shape, size and location) to diagnose herniation from 3D MR and CT axial (transverse sections) volumes of the discs. They also discussed the diagnosis of 16 clinical cases of various lumbar herniation types and report the follow-up period for 1.8 years. 75% of the patients showed excellent outcome after the surgery based on thier diagnosis while the rest 25% ranged between good and no improvement.

3. PROPOSED MODEL

Because intensity is the main factor in diagnosis of detection, we capture desiccation n_i with a Gibbs model:

$$P(n_i|d_i, \sigma_{\mathtt{I}(d_i)}) = \frac{1}{Z[n_i]} \exp^{-\mathtt{E}_{n_i}(d_i, \sigma_{\mathtt{I}(d_i)})} \quad (2)$$

where n_i is a binary random variable for desiccation of the disc i and $n_i \in \mathcal{N} = \{n_i : 1 \leq i \leq 6\}$; d_i is the location of the disc i and $d_i \in \mathcal{D} = \{d_i : 1 \leq i \leq 6\}$; σ_{d_i} is a neighborhood of pixels around the disc location d_i. $\mathtt{E}_{n_i}(d_i, \sigma_{\mathtt{I}(d_i)})$ is the energy function identified by disc location d_i and the intensity of a pixel neighborhood $\sigma_{\mathtt{I}(d_i)}$.

We use two potentials that represent the appearance $\mathtt{I}$ and the context in intensity between discs $(i \sim j)$. Our energy function $\mathtt{E}_{n_i}(d_i, \sigma_{\mathtt{I}(d_i)}))$ is:

$$\mathtt{E}_{n_i}(d_i, \sigma_{\mathtt{I}(d_i)}) = \Big[\beta_1 \sum_{d \in \mathcal{D}} U_{\mathrm{I}}(d_i, \sigma_{\mathtt{I}(d_i)}) + \beta_2 \sum_{(i \sim j)} V_{\mathrm{D}}(\widetilde{\sigma_{\mathtt{I}(d_i)}}, \widetilde{\sigma_{\mathtt{I}(d_j)}}; n_i, n_j)\Big] \quad (3)$$

where β_1 and β_2 are the model parameters that control the effect of appearance and context on the inference. $\widetilde{\sigma_{\mathtt{I}(d_i)}}$ is the median intensity level for the disc pixel neighborhood $\sigma_{\mathtt{I}(d_i)}$. U_{I} is the appearance potential which is a model of disc intensity neighborhood surrounding each disc $d_i \in \mathcal{D}$ and the intensity of the pixel neighborhood $\sigma_{\mathtt{I}(d_i)}$ of that location. V_{D} is the context potential which we model as a Bayesian model-aware affinity (Corso et al. [15]) that handles context $\mathcal{D}$ between neighboring discs $(i \sim j)$.

Our model requires two inputs: the locations of the discs $\mathcal{D} = \{d_1, d_2, ..., d_6\}$, and the intensity of a neighborhood surrounding every location $\sigma_{I(d_i)}$. The first input is actually the outcome of the labeling problem which we produce from our previous work [9]. The second input is obtained from the image intensity $\mathtt{I} = \{\text{ Intensity} : 0 \leq \text{Intensity} \leq 2^b - 1\}$ for the disc location and a defined neighborhood σ_{d_i} where b is the bit depth of the images, which is 12 bits for our dataset.

It is very important to normalize the intensity values $\mathtt{I}$ of the image based on the intensity of the spine so that we have a standard reference for the intensity [8] as discussed earlier in Section 1. Below, we discuss the model for the two potentials:

Appearance potential $U_{\mathtt{I}}(d_i, \sigma_{\mathtt{I}(d_i)})$ models the expected intensity level of the desiccated discs, which we model as Gaussian. After taking the negative log:

$$U_{\mathtt{I}}(d_i, \sigma_{\mathtt{I}(d_i)}) = \frac{\sum_{j \in \sigma_{\mathtt{I}(d_i)}} (\mathtt{I}(j) - \mu_{\mathtt{I}})^2}{2\sigma_{\mathtt{I}}^2} \quad (4)$$

where d_i is the location $d_i = (row, col)$ of disc i, $\mathtt{I}(d_i)$ is the intensity at location d_i, σ_{d_i} is some pixel neighborhood of the location d_i, $\mu_{\mathtt{I}}$ is the expected intensity levels of the desiccated discs, $\sigma_{\mathtt{I}}^2$ is the variance of the intensity levels of desiccated discs. Both $\mu_{\mathtt{I}}$ and $\sigma_{\mathtt{I}}^2$ are learned from the training data where a set of images are manually labeled (or labeled by our labeling method in [9]).

Context potential $V_{\mathrm{D}}(\widetilde{\sigma_{\mathtt{I}(d_i)}}, \widetilde{\sigma_{\mathtt{I}(d_j)}}; n_i, n_j)$ models the contextual relationship between neighboring disc locations i and j. We are looking for evaluation of disc i along with its context. However, because the labels of this context $\{j : 1 \leq j \leq 6\}$ and $\{j \neq i\}$ are unknown, we marginalize over all possibilities of this context which gives us a Bayesian estimate of the contexual energy based on the model-aware affinity, which is similar to Corso et al. [15]. Thus we have:

$$V_{\mathrm{D}}(\widetilde{\sigma_{\mathtt{I}(d_i)}}, \widetilde{\sigma_{\mathtt{I}(d_j)}}; n_i, n_j) = \sum_{n_j} \hat{V}_{\mathrm{D}}(\widetilde{\sigma_{\mathtt{I}(d_i)}}, \widetilde{\sigma_{\mathtt{I}(d_j)}}; \quad n_i, n_j) P(n_j | \sigma_{\mathtt{I}(d_j)}, \mathtt{I}) \quad (5)$$

where $P(n_j|\sigma_{\mathtt{I}(d_i)}, \mathtt{I})$ is the intensity model $U_{\mathtt{I}}$ for disc j, and $\hat{V}_{\mathrm{D}}(\widetilde{\sigma_{\mathtt{I}(d_i)}}, \widetilde{\sigma_{\mathtt{I}(d_j)}}; n_i, n_j)$ is a Gaussian model for the difference in intensity between the medians:

$$\hat{V}_{\mathrm{D}}(\widetilde{\sigma_{\mathtt{I}(d_i)}}, \widetilde{\sigma_{\mathtt{I}(d_j)}}; n_i, n_j) = \frac{(q_{ij} - \mu_Q)^2}{\sigma_Q^2} \quad (6)$$

where $\widetilde{\sigma_{\mathtt{I}(d_i)}}$, and $\widetilde{\sigma_{\mathtt{I}(d_j)}}$ are the median intensity values of the pixel neighborhood of discs i and j, respectively. μ_Q is the expected difference between median intensities and σ_Q^2 is the variance of differences between median intensities. We learn both μ_Q and σ_Q^2 from the training data. We define q_{ij} as:

$$q_{ij} = |\widetilde{\sigma_{\mathtt{I}(d_j)}} - \widetilde{\sigma_{\mathtt{I}(d_j)}}| \quad (7)$$

4. EXPERIMENTAL RESULTS AND DATA

We use a dataset of 55 clinical MRI volumes containing normal and desiccated cases. However, other abnormalities exist such as herniation and spinal stenosis. We use the T2-weighted volumes for training and testing our proposed model because T2-weighted have shown the highest sensitivity for disc signal [2, 3]. We pick the middle slice from every volume to represent that case and use it in our model training and testing. In desiccation, the middle slice is enough in diagnosis unlike other abnormalities such as herniation which requires the whole sagital volume or at least three slices beside the axial view in most cases. The diagnosis of desiccation requires detection of relative intensity of discs compared to the standard clincal signal as we discussed in section 1.

We perform ground truth annotation for our dataset by selecting a point inside every disc that roughly represents the center for that disc d_i and determine whether the disc is desiccated or normal.

It is worth mentioning that inter-observer error exist in lumbar diagnosis similar to many diagnosis tasks from various imaging modalities including plain radiographs, MRI, CT, SPECT (single-photon emission computed tomography), and High Resolution (HR). However, MRI shows high inter-observer reliability compared to plain radiographs in lumbar area diagnosis (e.g., [16]). Mulconrey et al. [17] showed that abnormality detection for degenerative disc and spondylolisthesis with MRI has $\kappa = 0.773$ and $\kappa = 0.728$, respectively, which is considered high in showing inter-observer reliability where this reliability is considered perfect when $0.8 \leq \kappa \leq 1$.

We perfom a cross-validation experiment using the 55 cases to train and test our proposed method. In every round, we separate thirty cases and train on the rest 25 cases. We perfom 10 rounds and every time the cases are selected randomly. We define the accuracy by:

$$Accuracy_i = 1 - \frac{1}{K} \sum_{j=1}^{K} |g_{ij} - n_{ij}| * 100\% \quad (8)$$

where $Accuracy_i$ represents the classification accuracy at the lumbar disc level i where $1 \leq i \leq 6$, the value K represents the number of cases in every experiment, g_{ij} is the ground truth binary assignment for disc i, and n_{ij} is the resulting binary assignment for disc i from the inference on our model. g_i and n_i are assigned the binary values such that they get the value 1 if i is a normal disc and 2 if it is a desiccated disc.

It is worth mentioning that we measure accuracy at every lumbar disc level separately to show the detailed classification accuracy at every level and thus have more understanding of the disc levels and its influence on classification accuracy. This appears in the row before the last in Table 1 where every value is a percentage accuarcy that represents the average of all the rounds in the experiment for every disc level. At the same time, we report the average accuracy for all the discs together for each round which is the last column in Table 1. The overall average accuracy for discs for all rounds in the experiment appears in the bottom-right cell in the same table.

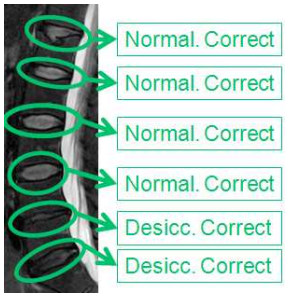
As shown in Table 1, the lower disc levels $E6, E5$, and $E4$ have less classification accuracy than the upper three levels. We believe that it is due to the fact that lower lumbar discs are usually more affected by abnormalities than the

upper ones. There are more desiccated lower lumbar discs than the upper ones in general. In our dataset, over 90% of desiccated discs are among the lower three discs.

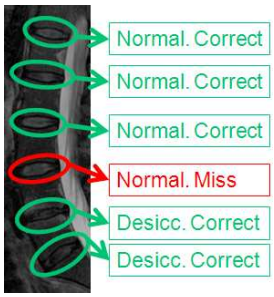
Fig. 2 shows two sample cases of classification output from inferencing on our model. Fig. 2(a) shows a case where the two lower discs are desiccated and the rest are normal. All discs are correctly classified. Fig. 2(b) shows another case where the lower discs are desiccated and the rest are normal. The level $\mathcal{L}3-\mathcal{L}4$ is misclassified (false positive) as desiccated while its ground truth is normal. The signal level of this disc indicates a start of desiccation as it is lower than the upper discs. All the rest discs are correclty classified.

Table 1. Classification results for the cross-validation experiment on 55 cases. Row before last shows average accuracy at every lumbar disc level and the last column shows the average accuracy for every round of 30 cases. We achieve an average of over 96% of classification accuracy.

Set	E6	E5	E4	E3	E2	E1	Accuracy
1	29	28	28	29	30	30	96.7%
2	29	29	28	29	30	28	96.1%
3	29	28	28	30	28	30	96.1%
4	30	28	27	29	28	30	95.6%
5	30	27	28	30	30	29	96.7%
6	30	28	28	29	30	30	97.2%
7	28	29	28	30	29	29	96.1%
8	28	29	28	29	30	30	96.7%
9	28	30	28	29	30	30	97.2%
10	29	28	28	29	30	29	96.1%
(%)	96.7	94.7	93.0	97.7	98.3	98.3	-
Average Accuracy							**96.4%**



(a) Desiccated discs: $\mathcal{L}4-\mathcal{L}5$ and $\mathcal{L}5-\mathcal{S}$. All levels are correclty classified.



(b) Desiccated levels: $\mathcal{L}4-\mathcal{L}5$ and $\mathcal{L}5-\mathcal{S}$. Level $\mathcal{L}3-\mathcal{L}4$ is misclassified (false positive - ground truth is Normal).

Fig. 2. Two sample cases: Green means it is correctly classified while red means otherwise.

5. CONCLUSION

We proposed a probabilistic model for inferencing on intensity levels of desiccated discs in clinical T2-weighted MRI for lumbar area. Our model successfully models the intensity levels of desiccated and normal discs. Besides, it models the intensity context using a Bayesian model-aware affinity that handles the neighborhood relation at object (disc) level. We achieve over 96% accuracy on cross-validation experiment on 55 clinical MRI cases that include desiccated discs and other abnormalities.

6. REFERENCES

[1] National Institute of Neurological Disorders and Stroke (NINDS), "Low back pain fact sheet," *NIND brichure*, 2008.

[2] Howard S. An and Paul A. Anderson et al., "Disc degeneration: summary," *Spine*, vol. 29, pp. 2677–2678, Dec. 2004.

[3] T. Videman and P. Nummi et al., "Digital assessment of mri for lumbar disc desiccation: A comparison of digital versus subjective assessments and digital intensity profiles versus discogram and macroanatomic findings," *Spine*, vol. 19, pp. 192–198, 1994.

[4] K. Luoma and R. Raininko et al., "Is the signal intensity of cerebrospinal fluid constant? intensity measurements with high and low field magnetic resonance imagers," *MRI J.*, vol. 11, pp. 549555, 1993.

[5] E. K. Luoma and R. Raininko et al., "Suitability of cerebrospinal fluid as a signal-intensity reference on mri: evaluation of signal-intensity variations in the lumbosacral dural sac," *Neuroradiology*, vol. 39, pp. 728–732, Oct. 1997.

[6] T. Videman and M.C. Batti et al., "Associations between back pain history and lumbar mri findings," *Spine*, vol. 28, pp. 582588, 2003.

[7] T. Videman and M.C. Batti et al., "Determinants of the progression in lumbar degeneration: a 5-year follow-up study of adult male monozygotic twins," *Spine*, vol. 31, pp. 671678, 2006.

[8] R. Niemelinen and T. Videman et al., "Quantitative measurement of intervertebral disc signal using mri," *Clin. Rad.*, vol. 63, no. 3, pp. 252 – 255, 2008.

[9] J. J. Corso, R. S. Alomari, and V. Chaudhary, "Lumbar disc localization and labeling with a probabilistic model on both pixel and object features.," in *Proc. of MICCAI*. 2008, vol. 5241 of *LNCS Part 1*, pp. 202–210, Springer.

[10] Richard S. Snell, *Clinical Anatomy by Regions*, Lippincott Williams and Wilkins, 8th edition, 2007.

[11] D. G. Bounds and P.J. Lloyd et al., "A multilayer perceptron network for the diagnosis of low back pain," San Diego, CA, July 1988, vol. 2, pp. 481–489.

[12] M. Vaughn, "Using an artificial neural network to assist orthopaedic surgeons in the diagnosis of low back pain," Department of Informatics, Cranfield University (RMCS), 2000.

[13] M. Tsai, S. Jou, and M. Hsieh, "A new method for lumbar herniated inter-vertebral disc diagnosis based on image analysis of transverse sections," *CMIG*, vol. 26, no. 6, pp. 369 – 380, 2002.

[14] J. P. Jenkins and D. S. Hickey et al., "Mr imaging of the intervertebral disc: A quantitative study," *Br. J. of Radiology*, vol. 58, no. 692, pp. 705–709, 1985.

[15] J. J. Corso and E. Sharon et al., "Efficient Multilevel Brain Tumor Segmentation with Integrated Bayesian Model Classification," *IEEE Trans. on Med. Imag.*, vol. 27, no. 5, pp. 629–640, 2008.

[16] S. S. Madan and M. Deanery, "Interobserver error in interpretation of the radiographs for degeneration of the lumbar spine," *Iowa Orthopaedic J.*, pp. 51–56, 2003.

[17] D . Mulconrey and R . Knight et al., "Interobserver reliability in the interpretation of diagnostic lumbar mri and nuclear imaging," *Spine*, vol. 6, pp. 177 – 184, 2006.