

# A Spatially Constrained Probabilistic Model for Robust Image Segmentation

Abhirup Banerjee and Pradipta Maji

**Abstract**—In general, the hidden Markov random field (HMRF) represents the class label distribution of an image in probabilistic model based segmentation. The class label distributions provided by existing HMRF models consider either the number of neighboring pixels with similar class labels or the spatial distance of neighboring pixels with dissimilar class labels. Also, this spatial information is only considered for estimation of class labels of the image pixels, while its contribution in parameter estimation is completely ignored. This, in turn, deteriorates the parameter estimation, resulting in sub-optimal segmentation performance. Moreover, the existing models assign equal weightage to the spatial information for class label estimation of all pixels throughout the image, which, create significant misclassification for the pixels in boundary region of image classes. In this regard, the paper develops a new clique potential function and a new class label distribution, incorporating the information of image class parameters. Unlike existing HMRF model based segmentation techniques, the proposed framework introduces a new scaling parameter that adaptively measures the contribution of spatial information for class label estimation of image pixels. The importance of the proposed framework is depicted by modifying the HMRF based segmentation methods. The advantage of proposed class label distribution is also demonstrated irrespective of the underlying intensity distributions. The comparative performance of the proposed and existing class label distributions in HMRF model is demonstrated both qualitatively and quantitatively for brain MR image segmentation, HEp-2 cell delineation, natural image and object segmentation.

**Index Terms**—Segmentation, expectation-maximization, clique potential, hidden Markov random field, class label distribution.

## I. INTRODUCTION

THE process of segmentation partitions an image into finite number of non-overlapping meaningful homogeneous regions. It is an invaluable step in many medical imaging applications. Brain image segmentation into cerebro-spinal fluid (CSF), gray matter (GM), and white matter (WM) tissues is a prerequisite step in several clinical investigations. For example, multiple sclerosis requires accurate measurement of

WM lesions for determining the optimum dose of medicine, while schizophrenia and epilepsy require volumetric classification of CSF, GM, and WM for identifying morphological distinctions between patients.

One major problem in image segmentation is uncertainty, caused by imprecision in computations and vagueness in class definitions. One popular technique in modeling this uncertainty is the probabilistic model, which generally applies the expectation-maximization (EM) algorithm for pixel classification according to the probability values, calculated from the image intensity distribution. Assuming a suitable probability distribution for image intensities, the segmentation methods aim to estimate the associated class label, based on pixel intensities. This estimation problem is usually formulated using maximum likelihood (ML) or maximum a posteriori (MAP) principles. In this regard, the finite mixture (FM), specifically, the finite Gaussian mixture (FGM) model, has become one of the most popular models for parametric model based image segmentation [1]–[5]. Wells et al. [1] used the EM algorithm in FGM model for optimal brain MR image segmentation, minimizing simultaneously the bias field artifacts. Alternatively, Greenspan et al. [3] used large number of Gaussian components to model each image class in FGM framework. Nguyen et al. [4] improved the FGM by varying the prior distribution of each image class for each pixel, depending on its neighbors. Liu and Zhang [5] integrated the level set approach into the FGM to achieve robust segmentation for noisy and bias field affected images.

Since none of the aforementioned works takes into consideration the spatial information for segmentation, the Markov random field (MRF) model has been introduced into the FM model based probabilistic frameworks [6]–[9]. Integrating the spatial information from neighboring pixels with their intensity distribution, Zhang et al. [6] introduced the HMRF model and proposed a joint EM-HMRF framework for robust image segmentation. Held et al. [7] proposed an MRF segmentation algorithm, using the adaptive segmentation approach of Wells et al. [1] over three important image features, namely, nonparametric image intensity distributions, neighborhood correlations, and signal inhomogeneities. Diplaros et al. [8] developed a generative framework, assuming the hidden class labels are generated using prior distributions of similar parameters from neighboring pixels. Nguyen and Wu [9] developed a novel technique using a simple metric to introduce spatial dependency among neighbors in joint FGM-MRF framework. Roche et al. [10] developed a new clique potential function using the distance weighting measure for HMRF model based segmentation in anisotropic images. Recently, Banerjee and

A. Banerjee is with the Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford, Oxford, United Kingdom and the Radcliffe Department of Medicine, Division of Cardiovascular Medicine, University of Oxford, Oxford, United Kingdom. E-mail: abhirup.banerjee@cardiov.ox.ac.uk.

P. Maji is with the Biomedical Imaging and Bioinformatics Lab, Machine Intelligence Unit, Indian Statistical Institute, Kolkata, India. E-mail: pmaji@isical.ac.in.

This work was done when A. Banerjee was a senior research fellow of Machine Intelligence Unit, Indian Statistical Institute, Kolkata, India. This work is partially an outcome of the R&D work undertaken by P. Maji in the project under the Visvesvaraya PhD Scheme of Ministry of Electronics and Information Technology, Government of India.

This paper has supplementary downloadable material available at <http://ieeexplore.ieee.org>, provided by the authors.

Maji [11] have introduced the stomped normal (StN) distribution for modelling brain tissues in MR images in rough-probabilistic framework. In another work [12], the image intensity distribution has been modelled using finite Student's- $t$  mixture model for brain MR image segmentation.

The MRF, more specifically the HMRF framework [6]–[12], generally models the class label distribution of an image by considering the spatial information from the neighboring pixels. In general, this spatial information consists of either number of neighboring pixels with similar class labels or the spatial distance of the neighboring pixels with dissimilar class labels. Also, this spatial information is only considered for class label estimation of image pixels, while its contribution in parameter estimation is completely ignored. This, in turn, may limit the class parameters from converging to their optimal estimates, thus resulting in sub-optimal segmentation. Another crucial drawback of the existing HMRF model based image segmentation techniques is that the models assume the spatial information is equally important to estimate the class labels of all pixels throughout the image. However, in general, only the pixels that lie completely inside an image class are properly classified by using this spatial information, while it creates ambiguity for classification of the pixels that lie at the boundary regions of image classes.

In this regard, the objective of the current research work is to present a new way to introduce the information of image class parameters for modeling the class label distribution of an image. By introducing a new clique potential function, the proposed framework presents a new class label distribution in HMRF based image segmentation. It enables the segmentation algorithm to optimally achieve the pixel labeling, as well as the estimates of the image class parameters. Also, introduction of a new adaptive scaling parameter, for estimating the importance of spatial information in class label estimation, facilitates the probabilistic framework to obtain improved classification accuracy, even for boundary pixels of image classes. The efficacy of the proposed approach is demonstrated by modifying HMRF based image segmentation algorithms, where the image is represented as finite mixture of Gaussian, Student's- $t$ , or StN [11] distributions. The comparative performance of the proposed and existing class label distributions in HMRF framework is demonstrated quantitatively as well as qualitatively for real and simulated brain MR image segmentation, HEp-2 cell delineation, natural image and object segmentation.

## II. BASICS OF HMRF MODEL BASED SEGMENTATION

This section describes the basics of HMRF model and HMRF based segmentation, where the image is represented as finite mixture of Gaussian, Student's  $t$ , or StN distributions.

### A. Basics of Hidden Markov Random Field

Let  $\mathcal{X}$  and  $\mathcal{Y}$  be two random fields with state spaces  $\mathcal{L} = \{1, 2, \dots, L\}$  and  $\mathcal{D} = \{1, 2, \dots, D\}$ , respectively. Let  $x$  and  $y$  be observations of  $\mathcal{X}$  and  $\mathcal{Y}$ , respectively, and  $\mathcal{X}$  and  $\mathcal{Y}$  denote their respective sample spaces. Given  $X_i = l$ ,  $Y_i$  follows a conditional probability distribution

$$p(y_i|l) = f(y_i; \theta_l) \quad \forall l \in \mathcal{L} \quad (1)$$

where  $\theta_l$  denotes the parameter set of class with label  $l$ . We also assume that  $(\mathcal{X}, \mathcal{Y})$  is pairwise independent. Now, in FM model, it is assumed that

$$\omega_l = p(X_i = l) \quad \forall l \in \mathcal{L}, i \in \mathcal{S} \quad (2)$$

is independent of the individual sites  $i \in \mathcal{S} = \{1, 2, \dots, N\}$ . So, the marginal distribution of  $Y_i = y$ , depending on parameter set  $\theta = \{\theta_l = (\omega_l, \theta_l) : l \in \mathcal{L}\}$ , reduces to

$$p(y|\theta) = \sum_{l \in \mathcal{L}} p(X_i = l, Y_i = y|\theta_l) = \sum_{l \in \mathcal{L}} \omega_l f(y; \theta_l). \quad (3)$$

This model is defined as the FM model. Now, in an MRF, the elements in  $\mathcal{S}$  are connected with each other by a neighborhood system, defined as  $\mathcal{N} = \{\mathcal{N}_i : i \in \mathcal{S}\}$ , where  $\mathcal{N}_i$  denotes the set of elements neighboring  $i$ . A random field  $\mathcal{X}$  is called an MRF over a neighborhood system  $\mathcal{N}$  iff

$$(i) p(x) > 0 \quad \forall x \in \mathcal{X}; \quad \text{and} \quad (ii) p(x_i|x_{\mathcal{S}-\{i\}}) = p(x_i|x_{\mathcal{N}_i}).$$

In addition to the MRF assumptions, if

- 1) the state of  $\mathcal{X}$  is unobservable,
- 2) given any  $x \in \mathcal{X}$ , every  $Y_i$  follows a known conditional distribution  $p(y_i|x_i) = f(y_i; \theta_{x_i})$ , and
- 3) for any  $x \in \mathcal{X}$ ,  $Y_i$ 's are conditionally independent,

the model is called the hidden MRF (HMRF). Using these assumptions, the marginal distribution of  $Y_i = y_i$ , given the parameter set  $\theta$  and  $X_i$ 's neighborhood  $X_{\mathcal{N}_i}$ , is modified as

$$p(y_i|x_{\mathcal{N}_i}, \theta) = \sum_{l \in \mathcal{L}} p(l|x_{\mathcal{N}_i}) f(y_i; \theta_l). \quad (4)$$

According to Hammersley-Clifford theorem [13] for modeling an MRF, the random field  $\mathcal{X}$  can be described using the Gibbs distribution, as follows:

$$p(x) = \frac{1}{Z} \exp(-E(x)), \quad (5)$$

$Z$  being a normalizing constant, called the partition function. Here,  $E(x)$  is an energy function of the form

$$E(x) = \sum_{c \in \mathcal{C}} E_c(x), \quad (6)$$

which is a sum of clique potentials  $E_c(x)$  over all possible cliques  $\mathcal{C}$ . In HMRF, a clique  $c$  is defined as a subset of elements in  $\mathcal{S}$ , where each pair of distinct elements is neighbors. Accordingly,  $p(l|x_{\mathcal{N}_i})$  is derived as

$$p(l|x_{\mathcal{N}_i}) = \frac{1}{Z} \exp\left(-\sum_{j \in \mathcal{N}_i} E_c(l, x_j)\right). \quad (7)$$

In [6], Zhang et al. defined the clique potential function as

$$E_c(x_i, x_j) = -\delta(x_i - x_j), \quad (8)$$

$\delta(\cdot)$  being the Kronecker's delta function. In later works [11], [12], [14], [15], the clique potential function is modified as

$$E_c(x_i, x_j) = -a\delta(x_i - x_j), \quad (9)$$

where  $a$  is a scaling parameter. In [10], Roche et al. defined the clique potential function for anisotropic images, using the distance weighting measure, as follows:

$$E_c(x_i, x_j) = \frac{0.1}{d_{ij}} [1 - \delta(x_i - x_j)], \quad (10)$$

where  $d_{ij}$  is the Euclidean distance between pixels  $i$  and  $j$ .

### B. GHMRF: Gaussian Distribution for HMRF Model Based Segmentation

In [6], Zhang et al. introduced the HMRF into a finite mixture of Gaussian distributions to develop the GHMRF algorithm for image segmentation. The algorithm follows the same model of (4), where  $Y_i|X_i = l \sim N(\mu_l, \sigma_l^2)$ , that is,

$$f_G(y_i; \theta_l) = \frac{1}{\sqrt{2\pi}\sigma_l} \exp\left(-\frac{(y_i - \mu_l)^2}{2\sigma_l^2}\right). \quad (11)$$

Here  $y_i$  and  $x_i$ , respectively, denote the intensity and class label of  $i$ th pixel in the image,  $i \in \mathcal{S}$ ,  $x_i \in \mathcal{L}$ . The estimation of parameters  $\theta_l, l \in \mathcal{L}$ , is performed using the iterative EM algorithm. The optimal labeling of the pixels is estimated using the MAP criterion and the iterated conditional modes (ICM) algorithm [16], as follows:

$$\hat{x}_i = \arg \min_{x_i} \left[ \frac{(y_i - \mu_{x_i})^2}{2\sigma_{x_i}^2} + \log \sigma_{x_i} + \sum_{j \in \mathcal{N}_i} E_c(x_i, x_j) \right]. \quad (12)$$

In case, clique potential function (9) is used, (12) reduces to

$$\hat{x}_i = \arg \min_{x_i} \left[ \frac{(y_i - \mu_{x_i})^2}{2\sigma_{x_i}^2} + \log \sigma_{x_i} - a\hat{n}_i(x_i) \right], \quad (13)$$

where  $\hat{n}_i(x_i)$  represents the number of neighbors of  $i$ th pixel with class label  $x_i$ . In case, clique potential function (10) is used, (12) is modified as

$$\hat{x}_i = \arg \min_{x_i} \left[ \frac{(y_i - \mu_{x_i})^2}{2\sigma_{x_i}^2} + \log \sigma_{x_i} + 0.1 \sum_{\substack{j \in \mathcal{N}_i \\ x_j \neq x_i}} \frac{1}{d_{ij}} \right]. \quad (14)$$

### C. tHMRF: Student's $t$ -Distribution for HMRF Model Based Segmentation

In [12], the intensity distribution in each class is modeled using Student's  $t$ -distribution for HMRF based segmentation. The image is represented as a finite  $t$ -mixture as in (4), where

$$f_t(y; \theta_l) = \frac{\Gamma\left(\frac{\nu_l+1}{2}\right)}{\sigma_l \sqrt{\pi\nu_l} \Gamma\left(\frac{\nu_l}{2}\right) \left(1 + \frac{d(y; \mu_l, \sigma_l)}{\nu_l}\right)^{\frac{(\nu_l+1)}{2}}}, \quad (15)$$

$\nu_l$  being the number of degrees of freedom, and  $d(y; \mu, \sigma) = \frac{(y-\mu)^2}{\sigma^2}$  is the squared Mahalanobis distance between scalars  $y$  and  $\mu$  with variance  $\sigma^2$ . Optimal labeling of the pixels is estimated similarly, using the MAP criterion, ICM algorithm, and clique potential function of (9), as follows:

$$\hat{x}_i = \arg \min_{x_i} [-\log f_t(y; \theta_{x_i}) - a\hat{n}_i(x_i)]. \quad (16)$$

The optimal labeling using clique potential of (10) is similarly estimated as

$$\hat{x}_i = \arg \min_{x_i} \left[ -\log f_t(y; \theta_{x_i}) + 0.1 \sum_{\substack{j \in \mathcal{N}_i \\ x_j \neq x_i}} \frac{1}{d_{ij}} \right]. \quad (17)$$

### D. StNHMRF: Stomped Normal Distribution for HMRF Model Based Segmentation

Incorporating the concepts of rough sets in probabilistic model, the StN distribution [11] models an image class  $\Omega$  as the union of two disjoint sets, namely, lower approximation region  $\underline{A}(\Omega)$  and boundary region  $B(\Omega)$ . The same model of (4) can be used to represent the intensity distribution, where

$$f_{StN}(y; \theta_l) = \frac{1}{D_l \sigma_l} \phi(z_{il}), z_{il} = \begin{cases} k_l, & \text{if } i \in \underline{A}(\Omega_l) \\ \frac{y_i - \mu_l}{\sigma_l}, & \text{if } i \in B(\Omega_l) \end{cases} \quad (18)$$

$D_l = 2(1 - \Phi(k_l) + k_l \phi(k_l))$ ,  $k_l$  denotes the width parameter of StN distribution, and  $\phi()$  and  $\Phi()$  denote the probability density and distribution functions of standard normal distribution, respectively. The lower approximation  $\underline{A}(\Omega_l)$  and boundary region  $B(\Omega_l)$  of  $\Omega_l$  are defined as follows:

$$\underline{A}(\Omega_l) = \left\{ i \in \mathcal{S} : \left| \frac{y_i - \mu_l}{\sigma_l} \right| < k_l \right\}; \quad (19)$$

$$B(\Omega_l) = \{ i \in \mathcal{S} : i \notin \underline{A}(\Omega_p) \quad \forall p \in \mathcal{L} \}. \quad (20)$$

The optimal labeling of each pixel is estimated similarly using the MAP criterion, ICM algorithm, and clique potential function of (9), as follows:

$$\hat{x}_i = \arg \min_{x_i} \left[ \frac{1}{2} z_{ix_i}^2 + \log \sigma_{x_i} - a\hat{n}_i(x_i) \right]. \quad (21)$$

In a similar way, the labels are optimally estimated using clique potential function of (10) as

$$\hat{x}_i = \arg \min_{x_i} \left[ \frac{1}{2} z_{ix_i}^2 + \log \sigma_{x_i} + 0.1 \sum_{\substack{j \in \mathcal{N}_i \\ x_j \neq x_i}} \frac{1}{d_{ij}} \right]. \quad (22)$$

## III. PROPOSED MODEL

This section introduces a novel class label distribution, which incorporates the information of class parameters for robust image segmentation. A new scaling parameter is also introduced to determine the contribution of spatial information in labeling a pixel.

### A. Proposed Class Label Distribution

According to the HMRF model, the class labels of the pixels in an image are assumed to follow Gibbs distribution, as in (5). As only neighboring pairs are being considered as cliques, (6) can be rewritten as

$$E(\mathcal{X}) = \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{N}_i} E_c(x_i, x_j). \quad (23)$$

Both clique potential functions in (9) and (10) consider only class label information of its neighboring pixels. The function of (10) additionally incorporates the spatial distance of the neighboring pixels from the center pixel and, in turn, acts as a simple high-pass filter. Using the clique potential function of (9), the prior probability  $p(l|x_{\mathcal{N}_i})$  from (7) is derived as

$$p(l|x_{\mathcal{N}_i}) = \frac{\exp(a\hat{n}_i(l))}{\sum_{m \in \mathcal{L}} \exp(a\hat{n}_i(m))}. \quad (24)$$

Similarly, using (10),  $p(l|x_{\mathcal{N}_i})$  is reduced to

$$p(l|x_{\mathcal{N}_i}) = \frac{\exp\left(0.1 \sum_{j \in \mathcal{N}_i; x_j \neq l} d_{ij}^{-1}\right)}{\sum_{m \in \mathcal{L}} \exp\left(0.1 \sum_{j \in \mathcal{N}_i; x_j \neq m} d_{ij}^{-1}\right)}. \quad (25)$$

From both (24) and (25), it is clearly visible that the use of existing clique potential functions in HMRF model incorporates either the number of pixels with same class label or the spatial distance of pixels with different class labels. The existing models completely ignore the contribution of image class parameters for modeling the class label distribution, which, in turn, can deviate the segmentation algorithms from attaining the optimal estimates of parameters, thus leading to sub-optimal segmentation results. Also, the scaling parameter  $a$  in the existing clique potential function remains constant for all pixels throughout the image. It indicates that the use of spatial information for estimating prior probabilities or class labels is important equally for all pixels throughout the image. However, this assumption clearly does not satisfy for the pixels that reside at the boundary of image classes, because this information actually creates ambiguity in classifying the pixel into one of the neighboring image classes. As a result, the overall segmentation performance may be degraded.

To solve the aforementioned problems of existing clique potential functions in HMRF model, in the proposed work, the clique potential function is defined as

$$E_c(x_i, x_j) = \frac{a_i}{2|\mathcal{N}_i|} \left[ (\mu_{x_i} - \mu_{x_j})^2 \left( \frac{1}{\sigma_{x_i}^2} + \frac{1}{\sigma_{x_j}^2} \right) - 1 \right], \quad (26)$$

where  $a_i \geq 0$ . Unlike the existing clique potential functions of (9) and (10), the proposed clique potential function integrates the information of image class parameters with the class label information of neighboring pixels. The scaling parameter  $a_i$ , in proposed model, varies for each pixel  $i$  and hence, can regulate the contribution of spatial information for each pixel during the estimation of class labels and image class parameters.

The expression  $(\mu_{x_i} - \mu_{x_j})^2$  in (26) incorporates the assumption of similarity between class labels of neighboring image pixels into the prior probability distribution and hence, aims to produce piecewise constant image classes. The term  $\left(\frac{1}{\sigma_{x_i}^2} + \frac{1}{\sigma_{x_j}^2}\right)$  in (26) normalizes the variability of both image classes from the clique potential function. The final expression is reduced by a constant term 1, so that the clique potential function can regulate the contribution of spatial information, even when the neighboring pixels are equally labelled. Without reduction by 1, the clique potential function would attain its minimum value of 0 at  $x_i = x_j$ , that is, when the  $i$ th and  $j$ th pixels are equally labelled. In this specific case, the function would nullify the effect of scaling parameter  $a_i$ . This, in turn, would interrupt the control of the contribution of spatial information, although, in this specific case, the contribution of spatial information should be very high. Hence, in the proposed clique potential function, the final expression is reduced by a constant term 1.

## B. Estimating the Scaling Parameter

The scaling parameter  $a_i$  controls the contribution of spatial information for estimating the class label of  $i$ th pixel,  $i \in \mathcal{S}$ . The higher value of  $a_i$  indicates higher contribution of spatial information, while its lower value indicates lower importance of the same for estimating the class label. In case  $a_i$  becomes 0, it indicates that the  $i$ th pixel is spatially independent, that is, its class label does not depend on its neighboring pixels. Hence, in that case, the class label of  $i$ th pixel should only be determined based on its intensity. In general, the pixels, lying inside the image classes, are mostly influenced by use of spatial information, while this information creates ambiguity in class label estimation of pixels that reside at the boundary of image classes. Hence, from this perspective, the problem of estimating scaling parameter  $a_i$  boils down to determining whether a pixel lies inside or at the boundary of a class.

In general, the pixels in an image with similar intensities are spatially connected to each other by a neighborhood system. So, if most of the neighboring pixels of a specific pixel belong to the same class, it is extremely likely that the pixel is an inlier pixel (lies inside) of that class. Hence, the contribution of spatial information in estimating the class label of that pixel should be very high. The aforementioned criterion is easily evaluated by measuring the number of neighboring pixels that belong to each of the  $L$  image classes. Let us assume that among the neighboring pixels of  $i$ th pixel,  $n_{ip}$  number of pixels belong to the  $p$ th class  $\Omega_p$ , where  $p \in \mathcal{L} = \{1, 2, \dots, L\}$ . The corresponding relative frequency for each class is denoted by  $r_{ip}$ , where  $r_{ip} = \frac{n_{ip}}{|\mathcal{N}_i|}$ ,  $\forall p \in \mathcal{L}$ . The relative frequency values are then sorted into  $r_{i(p)}$ ;  $p \in \mathcal{L}$ , where  $r_{i(p)}$  represents the  $p$ th highest relative frequency of the occurrence of an image class in the neighborhood of pixel  $i$ . Let us denote this class by  $\Omega_{(p)}$ . In this regard, it should be noted that  $((1), (2), \dots, (L))$  is a permutation of  $\mathcal{L} = \{1, 2, \dots, L\}$ .

Now, if the highest relative frequency corresponding to pixel  $i$  exceeds the second highest relative frequency by a threshold  $T$ , that is, if  $(r_{i(1)} - r_{i(2)}) > T$ , it implies that at least 100T% neighboring pixels of pixel  $i$  belong to one image class, denoted by  $\Omega_{(1)}$ . For relatively higher value of  $T$ , the aforementioned criterion ensures that most of the neighboring pixels belong to a single image class, which, in turn, provides significant evidence that the pixel of interest lies inside that class. In the proposed model, the value of  $T$  is fixed to 0.50. So, to satisfy the aforementioned condition in 8-neighborhood system of an image, at least 6 pixels among 8 neighboring pixels must belong to one image class, while the other image classes can contain at most one of the remaining pixels. Now, in case pixel  $i$  lies at the boundary of an image class, there should exist at least two classes  $\Omega_l$  and  $\Omega_m$ ,  $l, m \in \mathcal{L}$ , where  $2 \leq n_{il}, n_{im} \leq 6$  (Fig. 1). Hence, if the proposed criterion is satisfied for pixel  $i$ , the pixel of interest can never be a boundary pixel and should definitely be an inlier pixel. So, in this case, spatial information should play key role in class label estimation of the pixel and hence, the value of scaling parameter  $a_i$  should be high for that pixel. In the proposed model, the value of  $a_i$  is estimated as  $a > 0$  in this case.

If the proposed condition does not hold, that is,  $(r_{i(1)} -$



Fig. 1. (Left) Minimum number of boundary pixels in 8-neighborhood. (Right) Different levels of neighborhood system in distinct colors.

$r_{i(2)} \leq T$ , then there does not exist sufficient evidence that most of the neighboring pixels belong to one image class and there is high possibility that the  $i$ th pixel lies at the boundary between two classes. So, inclusion of spatial information may lead to misclassification of the pixel in this case. Hence, the proposed model estimates the scaling parameter  $a_i$  to be 0 in this case, to nullify the effects of spatial information.

The aforementioned criterion can be extended to include different levels of neighboring pixels (Fig. 1). The  $k$ th order neighborhood system of pixel  $i$  is defined as

$$\mathcal{N}_i^{(k)} = \left\{ j = (j_x, j_y) : |j_x - i_x| \leq k, |j_y - i_y| \leq k, j \notin \mathcal{N}_i^{(k-1)} \right\}, \quad (27)$$

where  $i = (i_x, i_y) \in \mathcal{S}$  and  $k \geq 1$ . Clearly,  $\mathcal{N}_i^{(1)} = \mathcal{N}_i \quad \forall i \in \mathcal{S}$ . Different levels of neighboring pixels are assigned different weights according to their distance from center pixel. The weights for the  $k$ th order neighboring pixels are defined as

$$w_i^{(k)} = \frac{1}{2^k}. \quad (28)$$

In case, the spatial information is considered till the  $P$ th order neighboring pixels, the weights for the neighboring pixels in  $P$ th level will be doubled, that is,

$$w_i^{(k)} = \begin{cases} \frac{1}{2^k}, & \text{if } k < P \\ \frac{1}{2^{k-1}}, & \text{if } k = P. \end{cases} \quad (29)$$

The composite relative frequencies corresponding to  $i$ th pixel, in the  $k$ th order neighborhood system, are defined as follows:

$$r_{il}^{(k)} = \sum_{j=1}^k w_i^{(j)} \frac{n_{il}^{(j)}}{|\mathcal{N}_i^{(j)}|}; \forall l \in \mathcal{L}. \quad (30)$$

Clearly,  $r_{il}^{(1)} = r_{il}; \forall l \in \mathcal{L}$ . In the proposed method, up to second order neighboring pixels are only considered, that is,  $P = 2$ . So, in addition to  $(r_{i(1)}^{(1)} - r_{i(2)}^{(1)}) > T$ , if  $(r_{i(1)}^{(2)} - r_{i(2)}^{(2)}) > T$ , then most of both first and second order neighboring pixels of pixel  $i$  belong to  $\Omega_{(1)}$ , and hence the possibility of the pixel's residing inside the image class becomes very high. So, in this case, the contribution of spatial information in estimating the class label of the  $i$ th pixel should be very high. Hence, the proposed model estimates the scaling parameter  $a_i$  to be  $2a$  in this case. In case the criterion  $(r_{i(1)}^{(2)} - r_{i(2)}^{(2)}) > T$  is not satisfied, then also the pixel should be an inlier pixel, since  $(r_{i(1)}^{(1)} - r_{i(2)}^{(1)}) > T$ . However, since the pixel lies very near to the boundary region, the scaling parameter is assigned less weight and is estimated as  $a$ . Moreover, in case  $(r_{i(1)}^{(1)} - r_{i(2)}^{(1)}) \leq T$  and  $(r_{i(1)}^{(2)} - r_{i(2)}^{(2)}) > T$ , then also the pixel has high possibility to be an inlier pixel due to the possible presence of high number of noise and

outliers in first-order neighborhood. So, incorporating all the aforementioned results, the scaling parameter  $a_i$  for the  $i$ th pixel is estimated as

$$a_i = \begin{cases} 2a, & \text{if } (r_{i(1)}^{(2)} - r_{i(2)}^{(2)}) > T \text{ and } (r_{i(1)}^{(1)} - r_{i(2)}^{(1)}) > T \\ a, & \text{if } (r_{i(1)}^{(2)} - r_{i(2)}^{(2)}) > T \text{ and } (r_{i(1)}^{(1)} - r_{i(2)}^{(1)}) \leq T \\ a, & \text{if } (r_{i(1)}^{(2)} - r_{i(2)}^{(2)}) \leq T \text{ and } (r_{i(1)}^{(1)} - r_{i(2)}^{(1)}) > T \\ 0, & \text{otherwise.} \end{cases}$$

For any  $P$ th order neighborhood system, the estimation of scaling parameters can be extended as follows:

$$a_i = \begin{cases} Pa, & \text{if } (r_{i(1)}^{(j)} - r_{i(2)}^{(j)}) > T \quad \forall j = 1, 2, \dots, P \\ pa, & \text{if for any } p \in \{1, 2, \dots, P-1\}, \\ & \exists t_1, t_2, \dots, t_p \in \{1, 2, \dots, P\} \\ & \ni (r_{i(1)}^{(t_j)} - r_{i(2)}^{(t_j)}) > T \quad \forall j = 1, 2, \dots, p \\ 0, & \text{otherwise.} \end{cases}$$

### C. GHMRF\_S: GHMRF with New Class Label Distribution

This section incorporates the concept of proposed class label distribution into the GHMRF algorithm [6]. Let  $y_i$  and  $x_i$ , respectively, denote the intensity and class label of  $i$ th pixel,  $i \in \mathcal{S}$  and  $x_i \in \mathcal{L} = \{1, 2, \dots, L\}$ . The image is represented as a finite Gaussian mixture, as in (4), where  $f(y_i; \theta_l)$  follows from (11) and the prior function  $p(l|x_{\mathcal{N}_i})$  is derived using the proposed clique potential function defined in (26), as follows:

$$p(l|x_{\mathcal{N}_i}) = \frac{\exp\left(-\frac{a_i}{2|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} (\mu_l - \mu_{x_j})^2 \left(\frac{1}{\sigma_l^2} + \frac{1}{\sigma_{x_j}^2}\right)\right)}{\sum_{m \in \mathcal{L}} \exp\left(-\frac{a_i}{2|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} (\mu_m - \mu_{x_j})^2 \left(\frac{1}{\sigma_m^2} + \frac{1}{\sigma_{x_j}^2}\right)\right)}. \quad (31)$$

Assumption of statistical independence between pixel intensities given their class labels reduces the probability density of overall image to

$$p(\underline{y}|x_{\mathcal{N}}, \theta) = \prod_{i \in \mathcal{S}} p(y_i|x_{\mathcal{N}_i}, \theta) = \prod_{i \in \mathcal{S}} \sum_{l \in \mathcal{L}} p(l|x_{\mathcal{N}_i}) p(y_i|l). \quad (32)$$

To estimate parameters  $\theta = \{\mu_l, \sigma_l : l \in \mathcal{L}\}$  from above expression, the iterative EM algorithm is applied. In the proposed GHMRF\_S algorithm, the latent membership values  $\tau_{il}$  are estimated in the E-step, as follows:

$$\tau_{il}^{(t)} = \frac{p(l|x_{\mathcal{N}_i}) f_G(y_i; \theta_l^{(t)})}{\sum_{m \in \mathcal{L}} p(m|x_{\mathcal{N}_i}) f_G(y_i; \theta_m^{(t)})}, \quad (33)$$

where the prior probability  $p(l|x_{\mathcal{N}_i})$  follows the proposed class label distribution of (31). For parameter estimation in M-step, the  $Q$ -function or expected complete data log-likelihood is generated, as follows:

$$\begin{aligned} Q(\theta|\theta^{(t)}) &= \sum_{i \in \mathcal{S}} \sum_{l \in \mathcal{L}} \tau_{il}^{(t)} [\log p(l|x_{\mathcal{N}_i}) + \log p(y_i|l)] \\ &= \sum_{i \in \mathcal{S}} \sum_{l \in \mathcal{L}} \tau_{il}^{(t)} \left[ -\frac{a_i}{2|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} (\mu_l - \mu_{x_j})^2 \left(\frac{1}{\sigma_l^2} + \frac{1}{\sigma_{x_j}^2}\right) + \frac{a_i}{2} \right. \\ &\quad \left. - \frac{1}{2} \log 2\pi - \log \sigma_l - \frac{(y_i - \mu_l)^2}{2\sigma_l^2} \right]. \end{aligned}$$

Optimizing the  $Q$ -function, the optimal estimates of the parameters  $\mu_l$  and  $\sigma_l^2$  are obtained as

$$\hat{\mu}_l^{(t+1)} = \frac{\sum_{i \in S} \tau_{il}^{(t)} \left( y_i + \frac{a_i}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} \left( 1 + \frac{\sigma_l^2(t)}{\sigma_{x_j}^2(t)} \right) \mu_{x_j}^{(t)} \right)}{\sum_{i \in S} \tau_{il}^{(t)} \left( 1 + \frac{a_i}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} \left( 1 + \frac{\sigma_l^2(t)}{\sigma_{x_j}^2(t)} \right) \right)}; \quad (34)$$

$$\hat{\sigma}_l^{2(t+1)} = \frac{\sum_{i \in S} \tau_{il}^{(t)} \left( (y_i - \mu_l^{(t+1)})^2 + \frac{a_i}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} (\mu_l^{(t)} - \mu_{x_j}^{(t)})^2 \right)}{\sum_{i \in S} \tau_{il}^{(t)}}. \quad (35)$$

It is evident from the estimate of latent variables  $\tau_{il}$  and the parameters  $\mu_l$  and  $\sigma_l^2$  that the use of proposed clique potential function significantly modifies the existing GHMRF framework. Since the latent variables now take into consideration the effects of class parameters of neighboring pixels and, more importantly, the importance of spatial information via the prior probability  $p(l|x_{\mathcal{N}_i})$ , it improves the estimation of memberships of each pixel to each of the image classes and, in turn, improves the parameter estimation. In addition, the estimates (34) and (35) of  $\mu_l$  and  $\sigma_l^2$  in the proposed model are influenced by additional spatial information, which was not considered in the estimates of the same parameters in GHMRF algorithm. The additional spatial information considers the information of neighboring pixels based on the estimates of parameters from previous iteration and, in turn, regulates the parameters to converge to their most accurate and robust estimates, even in presence of outliers and heavy noise.

The optimal pixel labeling is estimated according to MAP criterion and ICM algorithm, as follows:

$$\begin{aligned} \hat{x}_i &= \arg \min_{x_i} \left[ \frac{(y_i - \mu_{x_i})^2}{2\sigma_{x_i}^2} + \log \sigma_{x_i} \right. \\ &\quad \left. + \frac{a_i}{2|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} (\mu_{x_i} - \mu_{x_j})^2 \left( \frac{1}{\sigma_{x_i}^2} + \frac{1}{\sigma_{x_j}^2} \right) \right] \\ &= \arg \min_{x_i} \left[ \frac{1}{2} d_{ix_i}^2 + \log \sigma_{x_i} + \frac{a_i}{2|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} A'_{x_i, x_j} \Sigma_{x_i, x_j}^{-1} A_{x_i, x_j} \right], \end{aligned} \quad (36)$$

where  $d_{ij}^2 = \frac{(y_i - \mu_j)^2}{\sigma_j^2}$ ,  $A_{ij} = (\mu_i - \mu_j \quad \mu_j - \mu_i)'$ , and  $\Sigma_{ij} = \text{diag}(\sigma_i^2, \sigma_j^2)$ . The variable  $d_{ix_i}$  denotes normalized Euclidean distance and the term  $\left[ \frac{1}{2} d_{ix_i}^2 + \log \sigma_{x_i} \right]$  represents the intensity information of the pixels, which is common in all three evaluation functions (36), (13), and (14). The three evaluation functions mainly differ in their definitions of class label distributions. The spatial information in both (13) and (14) relies, respectively, on the number of pixels with same class labels and the spatial distance from pixels with different class labels. Hence, this spatial information is not comparable with the pixel intensity information, measured in terms of a normalized distance between pixel intensities. In comparison, (36) represents the spatial information in terms of an average Mahalanobis distance between mean intensities of neighboring pixels and hence, is comparable with the pixel intensity information. Also, the new spatial information nullifies the effect of overall variation in each image class, using a normalized measure, and hence, can produce robust

estimates of class labels, compared to the GHMRF model. The incorporation of new adaptive scaling parameter  $a_i$  in the proposed model improves the labeling of pixels in boundary region of image classes. Similar modifications are observed in both  $t$ HMRF\_S and  $St$ NHMRF\_S algorithms.

#### D. $t$ HMRF\_S: $t$ HMRF with New Class Label Distribution

Similar to the previous model, in this case, the image is represented as a finite Student's  $t$ -mixture, as in (4), where  $f(y_i; \theta_l)$  follows from (15) and the prior  $p(l|x_{\mathcal{N}_i})$  is derived using the proposed clique potential function, as follows:

$$p(l|x_{\mathcal{N}_i}) = \frac{\exp \left( -\frac{a_i}{2|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} (\mu_l - \mu_{x_j})^2 \left( \frac{u_{il}}{\sigma_l^2} + \frac{u_{jx_j}}{\sigma_{x_j}^2} \right) \right)}{\sum_{m \in \mathcal{L}} \exp \left( -\frac{a_i}{2|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} (\mu_m - \mu_{x_j})^2 \left( \frac{u_{im}}{\sigma_m^2} + \frac{u_{jx_j}}{\sigma_{x_j}^2} \right) \right)}. \quad (37)$$

The latent variables  $u_{il}$  are estimated in the E-step as:

$$u_{il}^{(t)} = \frac{v_l^{(t)} + 1}{v_l^{(t)} + d(y_i; \mu_l^{(t)}, \sigma_l^{(t)})}, \quad (38)$$

and the  $\tau_{il}$  are estimated using (33), where the prior probability  $p(l|x_{\mathcal{N}_i})$  follows the proposed definition of (37). The parameters  $\mu_l$  and  $\sigma_l^2$  are estimated in the M-step, as follows:

$$\begin{aligned} \hat{\mu}_l^{(t+1)} &= \frac{\sum_{i \in S} \tau_{il}^{(t)} \left( u_{il}^{(t)} y_i + \frac{a_i}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} \left( u_{il}^{(t)} + u_{jx_j}^{(t)} \frac{\sigma_l^2(t)}{\sigma_{x_j}^2(t)} \right) \mu_{x_j}^{(t)} \right)}{\sum_{i \in S} \tau_{il}^{(t)} \left( u_{il}^{(t)} + \frac{a_i}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} \left( u_{il}^{(t)} + u_{jx_j}^{(t)} \frac{\sigma_l^2(t)}{\sigma_{x_j}^2(t)} \right) \right)}; \quad (39) \\ \hat{\sigma}_l^{2(t+1)} &= \frac{\sum_{i \in S} \tau_{il}^{(t)} u_{il}^{(t)} \left( (y_i - \mu_l^{(t+1)})^2 + \frac{a_i}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} (\mu_l^{(t)} - \mu_{x_j}^{(t)})^2 \right)}{\sum_{i \in S} \tau_{il}^{(t)} u_{il}^{(t)}}. \quad (40) \end{aligned}$$

$v_l^{(t+1)}$  is estimated numerically from

$$1 + \frac{\sum_{i \in S} \tau_{il}^{(t)} (\log u_{il}^{(t)} - u_{il}^{(t)})}{\sum_{i \in S} \tau_{il}^{(t)}} + \psi \left( \frac{v_l + 1}{2} \right) - \log \left( \frac{v_l + 1}{2} \right) + \log \left( \frac{v_l}{2} \right) - \psi \left( \frac{v_l}{2} \right) = 0, \quad (41)$$

where  $\psi(s) = \frac{1}{\Gamma(s)} \frac{\partial}{\partial s} \Gamma(s)$  represents the digamma function. Similar to the arguments presented in Section III-C, here also, the incorporation of new spatial information via prior probability  $p(l|x_{\mathcal{N}_i})$  improves the estimation of the latent variable  $\tau_{il}$  and in turn, improves the parameter estimation. Additionally, the incorporation of spatial information for parameter estimation, which was not considered in existing  $t$ HMRF model, regulates the parameters for accurate and robust estimation, even in heavy noisy environment. Although the estimation of  $v_l$  follows the same approach of  $t$ HMRF, the modified estimates of  $\tau_{il}$  in (41) improves its optimal estimation. The pixel labels are similarly measured using the MAP criterion and ICM algorithm, as follows:

$$\hat{x}_i = \arg \min_{x_i} \left[ -\log f_i(y; \theta_{x_i}) + \frac{a_i}{2|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} A'_{x_i, x_j} \Sigma_{x_i, x_j}^{-1} A_{x_i, x_j} \right]. \quad (42)$$

### E. *StNHMRFS*: *StNHMRF* with New Class Label Model

In this model, the image is represented as a finite *StN* mixture, as in (4), where  $f(y_i; \theta_l)$  follows the definition of (18) and the prior probability  $p(l|x_{\mathcal{N}_i})$  follows the definition of (31). The latent variables are estimated as

$$\tau_{il}^{(t)} = \begin{cases} 1, & \text{if } i \in \underline{A}(\Omega_l) \\ \frac{p(l|x_{\mathcal{N}_i}) f_{StN}(y; \theta_l^{(t)})}{\sum_{m \in \mathcal{L}} p(m|x_{\mathcal{N}_i}) f_{StN}(y; \theta_m^{(t)})}, & \text{else if } i \in B(\Omega_l) \\ 0, & \text{otherwise.} \end{cases} \quad (43)$$

The parameters  $\mu_l$  and  $\sigma_l^2$  are estimated in M-step, as follows:

$$\hat{\mu}_l^{(t+1)} = \alpha \mathcal{A}_l^{(t)} + (1 - \alpha) \mathcal{B}_l^{(t)}; \quad (44)$$

$$\text{where } \mathcal{A}_l^{(t)} = \frac{\sum_{i \in \underline{A}(\Omega_l)} \left( y_i + \frac{\alpha_i}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} \left( 1 + \frac{\sigma_l^2(t)}{\sigma_{x_j}^2(t)} \right) \mu_{x_j}^{(t)} \right)}{\sum_{i \in \underline{A}(\Omega_l)} \left( 1 + \frac{\alpha_i}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} \left( 1 + \frac{\sigma_l^2(t)}{\sigma_{x_j}^2(t)} \right) \right)},$$

$$\mathcal{B}_l^{(t)} = \frac{\sum_{i \in B(\Omega_l)} \tau_{il}^{(t)} \left( y_i + \frac{\alpha_i}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} \left( 1 + \frac{\sigma_l^2(t)}{\sigma_{x_j}^2(t)} \right) \mu_{x_j}^{(t)} \right)}{\sum_{i \in B(\Omega_l)} \tau_{il}^{(t)} \left( 1 + \frac{\alpha_i}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} \left( 1 + \frac{\sigma_l^2(t)}{\sigma_{x_j}^2(t)} \right) \right)},$$

$\alpha$  being the relative importance of lower approximation region;

$$\hat{\sigma}_l^{2(t+1)} = \frac{\sum_{i \in \mathcal{S}} \tau_{il}^{(t)} \left( (y_i - \mu_l^{(t+1)})^2 + \frac{\alpha_i}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} (\mu_{x_j}^{(t)} - \mu_{x_j}^{(t)})^2 \right)}{\sum_{i \in \mathcal{S}} \tau_{il}^{(t)}}. \quad (45)$$

Here also, the estimates of parameters differ from the existing *StNHMRF* algorithm in terms of the inclusion of spatial information of neighboring pixels. The estimates of the class labels using MAP criterion and ICM algorithm is given below:

$$\hat{x}_i = \arg \min_{x_i} \left[ \frac{1}{2} z_{ix_i}^2 + \log \sigma_{x_i} + \frac{\alpha_i}{2|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} A'_{x_i, x_j} \sum_{x_i, x_j}^{-1} A_{x_i, x_j} \right]. \quad (46)$$

## IV. EXPERIMENTAL RESULTS AND DISCUSSION

This section presents the comparative performance of the modified algorithms, termed as Algorithm\_S, over their two existing counter-parts, termed as Algorithm-1 and Algorithm-2, using clique potential functions (9) and (10), respectively. The performance of proposed *StNHMRFS* has been compared with state-of-the-art image segmentation algorithms: deviation-sparse FCM with neighbor information constraint (DSFCM\_N) [17], intuitionistic center-free FCM (ICFFCM) [18], improved FCM based on morphological reconstruction and membership filtering (FRFCM) [19], kernel metric and trade-off weighted fuzzy factor based fuzzy local information *c*-means (KWFLICM) [20], bias-corrected FCM (BCFCM) [21], multiplicative intrinsic component optimization (MICO) [22], adaptive segmentation (ASeg) [1], modified EM (mEM) [23], rough-fuzzy *c*-means (RFCM) [24], robust RFCM (rRFCM) [25]; and several analysis tools for MRI,

namely, statistical parameter mapping software (SPM) version 8 [26] and FMRIB Software Library (FSL) version 5.0 [27]. All the algorithms and measures are executed in Ubuntu 16.04 LTS 64-bit OS having machine configuration Intel(R) Core(TM) i7-2600 CPU @3.40GHz×8 and 16 GB RAM.

For analyzing the performance of different algorithms, the experiments are performed on some benchmark simulated MR images from ‘‘BrainWeb: Simulated Brain Database’’ (<http://www.bic.mni.mcgill.ca/brainweb/>) [28], [29], real MR images from ‘‘IBSR: Internet Brain Segmentation Repository’’ (<http://www.cma.mgh.harvard.edu/ibsr/>), indirect immunofluorescence (IIF) images obtained from ‘‘MIVIA HEp-2 Images Dataset’’ [30], ‘‘Berkeley Image Segmentation Dataset’’ [31], and ‘‘PASCAL Visual Object Classes (VOC) Dataset’’ [32]. The segmentation performance is evaluated using three quantitative indices, namely, Dice coefficient, sensitivity, and specificity. A good segmentation algorithm should make the values of these three indices as high as possible, and ideally, the values should be 1. The evaluation indices computed over different images are graphically presented using box plot. The significance analysis of the results are performed with the help of parametric paired *t*-test (one-tailed) and represented using the star (\*) symbol. The ‘\*’, ‘\*\*’, ‘\*\*\*’, and ‘\*\*\*\*’ symbols are used to indicate p-values of less than 0.05, less than 0.01, less than 0.001, and less than 0.0001, respectively; while the not significant ( $\geq 0.05$ ) p-values are marked as ‘n.s.’. Before applying the brain MR image segmentation algorithms, the brain extraction tool [33] is applied to remove the non-brain tissues like skull, scalp, dura, etc., from the images; while the RC2 bias field correction algorithm [34] is used to remove the bias field artifact present in MR images. The thresholding method due to Otsu [35] is incorporated for initial estimation of image class parameters. Detailed experimental results, along with brief description of the image databases, are provided in the Supplementary material.

### A. Importance of New Scaling Parameter

To establish the importance of new scaling parameter, the proposed parameter has been incorporated into both clique potential functions of (9) [6], [11], [14] and (10) [10], irrespective of the image intensity distributions (*StN*, *t*, and Gaussian). The efficacy of the scaling parameter has been tested on simulated and real brain MR images of BrainWeb and IBSR databases, HEp-2 cell IIF images of MIVIA database, and natural images of Berkeley and PASCAL VOC datasets. The quantitative results, in terms of box-plots, is presented in Fig. 2.

1) *StNHMRF Model*: From Fig. 2, it is observed that the use of new scaling parameter in clique potential function of (9) in *StNHMRF* model provides significant improvement in segmentation accuracy for brain MR image segmentation in BrainWeb and IBSR databases, natural image segmentation in Berkeley Dataset, and object segmentation in the PASCAL VOC dataset, irrespective of the evaluation indices. For HEp-2 cell segmentation in MIVIA database, the proposed scaling parameter also generates significant performance improvement for sensitivity and specificity indices; while for Dice coefficient, the results are better, but not statistically significant.

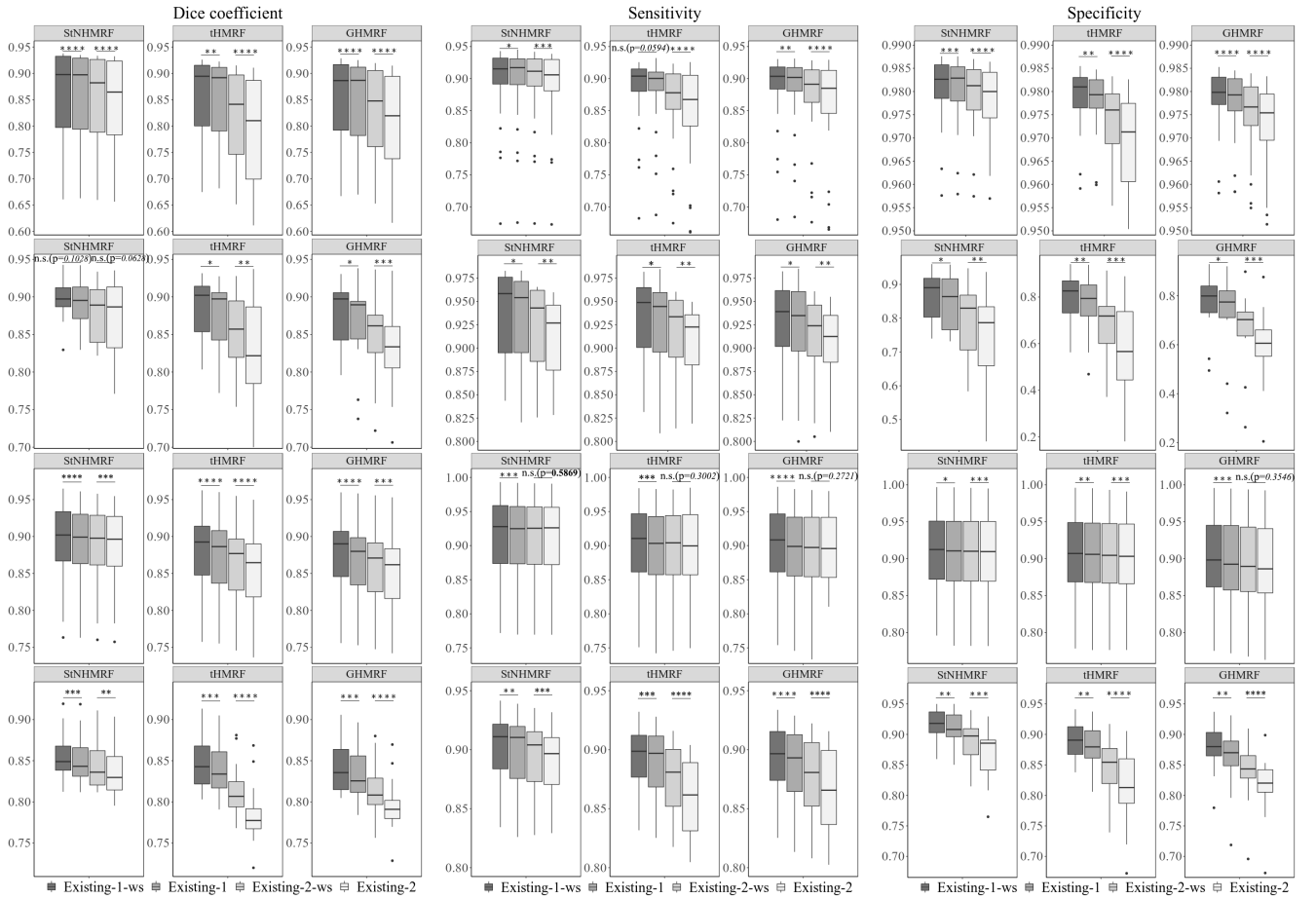


Fig. 2. Box plot depicting the importance of novel scaling parameter over two existing class label distributions ('ws' indicates 'with scaling parameter'): BrainWeb and IBSR databases (top); MIVIA database (second); Berkeley Segmentation Dataset (third); and PASCAL VOC dataset (last).

The use of new scaling parameter in clique potential function of (10) also provides significant improvement in segmentation accuracy for brain MR image segmentation in BrainWeb and IBSR databases and object segmentation in the PASCAL VOC dataset, irrespective of the segmentation evaluation indices. For HEP-2 cell segmentation, the proposed scaling parameter generates significant improvement for sensitivity and specificity indices, and better but not significant for Dice coefficient. For natural image segmentation in Berkeley Dataset, the proposed scaling parameter in (10) provides significantly better segmentation with respect to both Dice coefficient and specificity. However, with respect to sensitivity values, the existing approach provides better segmentation accuracy.

2) *tHMF Model*: The use of new scaling parameter in clique potential function of (9) in *tHMF* model provides significant improvement in segmentation accuracy for HEP-2 cell segmentation in MIVIA database, natural image segmentation in Berkeley Dataset, and object segmentation in the PASCAL VOC dataset. For brain MR image segmentation in BrainWeb and IBSR databases, the proposed scaling parameter also generates significant improvement for Dice coefficient and specificity indices; while for sensitivity, its results are better, but not statistically significant. For clique potential function of (10), the new scaling parameter also provides significant segmentation accuracy for brain MR image segmentation, HEP-2

cell segmentation, and object segmentation, irrespective of the evaluation indices. For natural image segmentation of Berkeley Dataset, the proposed scaling parameter generates significant improvement for both Dice coefficient and specificity, and better but not significant for sensitivity index.

3) *GHMRF Model*: The use of new scaling parameter in clique potential function of (9) in *GHMRF* model provides significant improvement in segmentation accuracy for all segmentation datasets, irrespective of the segmentation evaluation indices. For clique potential function of (10), the new scaling parameter also provides significant segmentation accuracy for brain MR image segmentation, HEP-2 cell segmentation, and object segmentation. For natural image segmentation in Berkeley Dataset, the proposed scaling parameter also generates significant performance improvement with respect to Dice coefficient and better but not statistically significant performance improvement with respect to sensitivity and specificity values.

### B. Importance of New Scaling Parameter in Novel Class Label Distribution

To establish the importance of new scaling parameter in proposed class label distribution, the experiments have been performed on all image databases, irrespective of the underlying image intensity models (*StNHMF*, *tHMF*, and



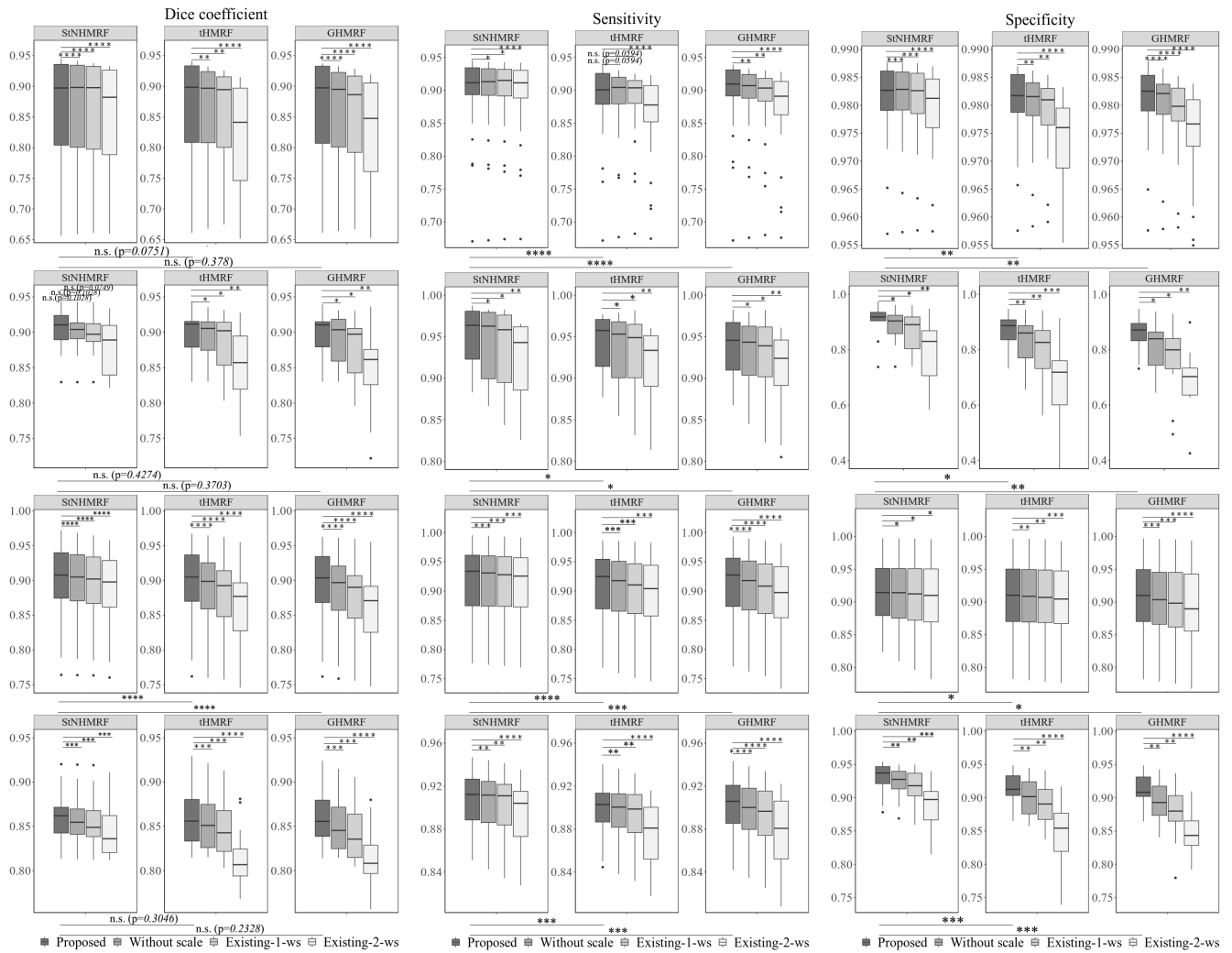


Fig. 3. Box plot depicting the importance of novel class label distribution: BrainWeb and IBSR databases (top); MIVIA database (second); Berkeley Segmentation Dataset (third); and PASCAL VOC dataset (last).

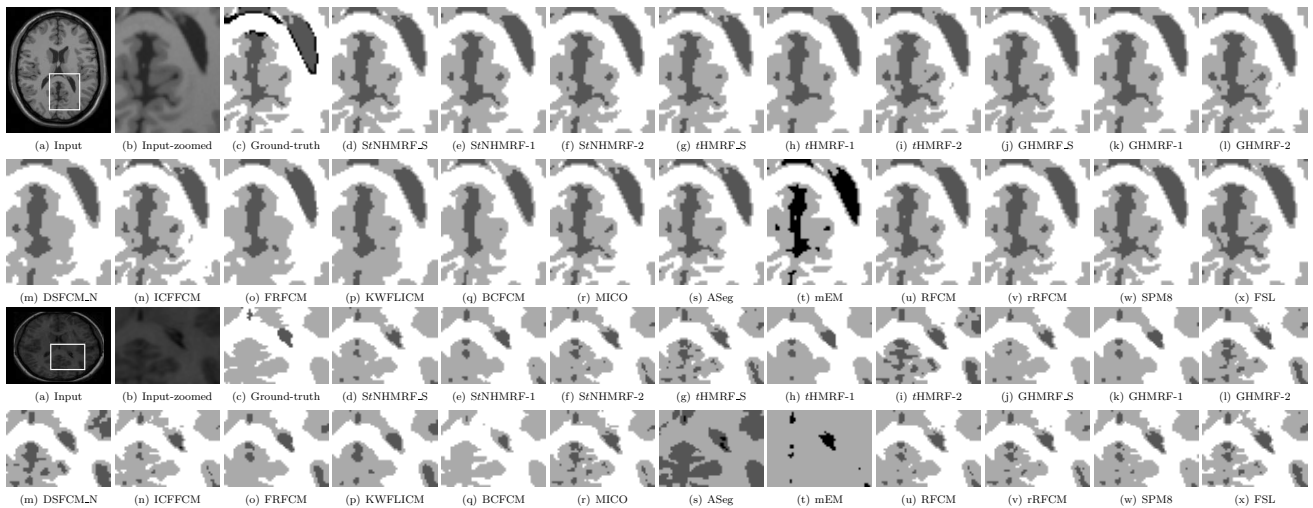


Fig. 4. Zoomed-in regions of segmented images obtained by different segmentation algorithms on BrainWeb database with 1% noise and 40% bias field (top two rows) and IBSR database volume number 3 (bottom two rows).

GHMRF). The quantitative results, in terms of box-plots, have been presented in Fig. 3. In *StNHMRF* model, the application

of new scaling parameter in proposed class label distribution has improved the segmentation accuracy significantly for brain

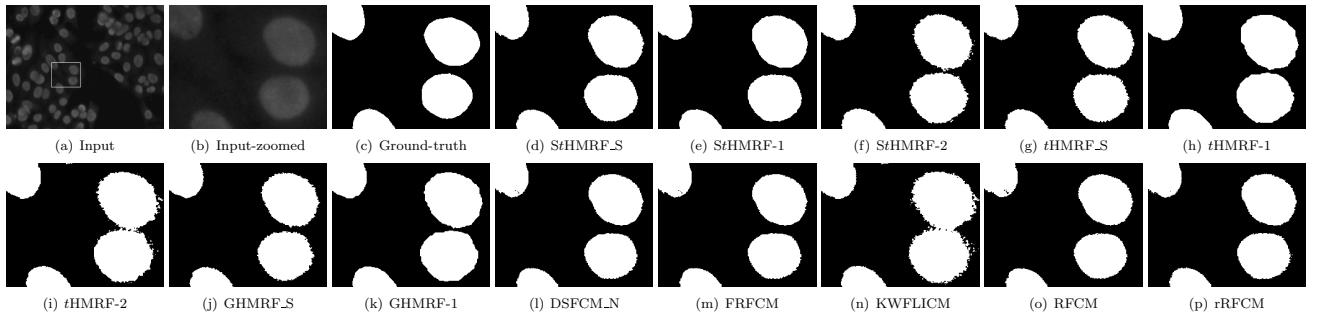


Fig. 5. Zoomed-in regions of segmented images obtained by different segmentation algorithms on MIVIA dataset image number 01.

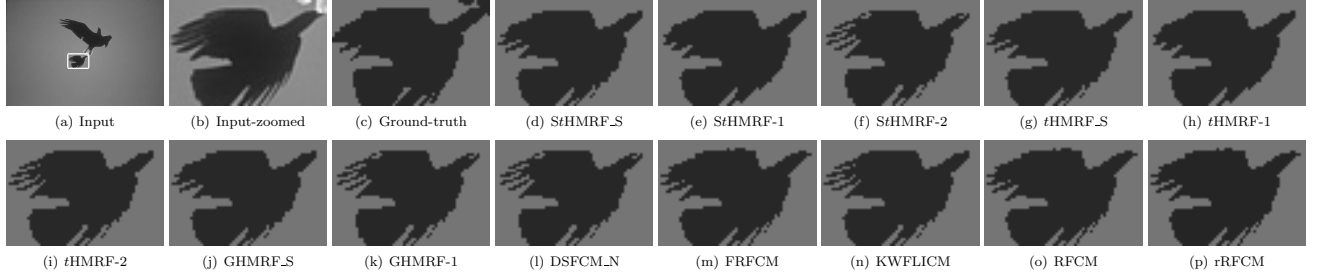


Fig. 6. Zoomed-in regions of segmented images obtained by different segmentation algorithms on Berkeley Segmentation Dataset image number 135069.

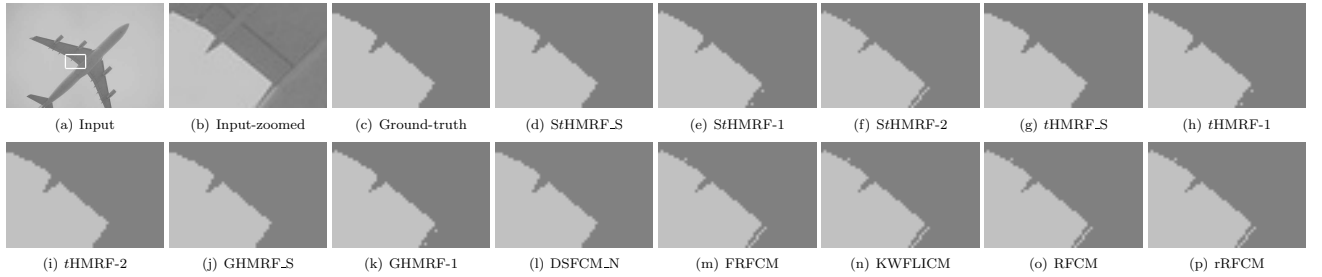


Fig. 7. Zoomed-in regions of segmented images obtained by different segmentation algorithms on PASCAL VOC dataset image number 2008\_005443.

MR image segmentation, natural image segmentation, and object segmentation. For HEP-2 cell segmentation, the proposed scaling parameter also generates significant performance improvement with respect to sensitivity and specificity values and better but not statistically significant improvement with respect to Dice coefficient. In  $t$ HMRF model, the application of new scaling parameter in proposed class label distribution has improved the segmentation accuracy significantly for HEP-2 cell segmentation, natural image segmentation, and object segmentation. For brain MR image segmentation, the proposed scaling parameter has also provided significant performance improvement with respect to Dice coefficient and specificity; while better but not statistically significant performance improvement is observed with respect to sensitivity values. In GHMRF model, the application of new scaling parameter has improved the segmentation accuracy significantly for proposed class label distribution, in all image segmentation databases.

### C. Importance of Novel Class Label Distribution

To establish the importance of novel class label distribution, the proposed class label distribution has been incorporated into

several existing HMRF based image segmentation algorithms and the results are depicted in Fig. 3 using box-plot.

1) *StNHMRF Model*: From the box-plots, it is observed that the use of new class label distribution in  $St$ NHMRF algorithm produces significantly better segmentation results for brain MR image segmentation, natural image segmentation, and object segmentation, than that obtained using both clique potential functions of (9) and (10) with new scaling parameter. For HEP-2 cell segmentation, the proposed class label distribution provides significant performance improvement with respect to sensitivity and specificity, and better but not statistically significant improvement with respect to Dice coefficient. Since the clique potential function of (9) considers only class label information of its neighboring pixels, its application in the HMRF model incorporates only the number of pixels with similar class label. This, in turn, puts higher weight to the spatial information compared to the pixel intensity information and hence, causes misclassification for pixels at the edge of tissue classes. The clique potential function of (10) additionally incorporates the spatial distance of the neighboring pixels from the center pixel and acts as a simple high-pass filter. Hence, the use of (10) in HMRF model

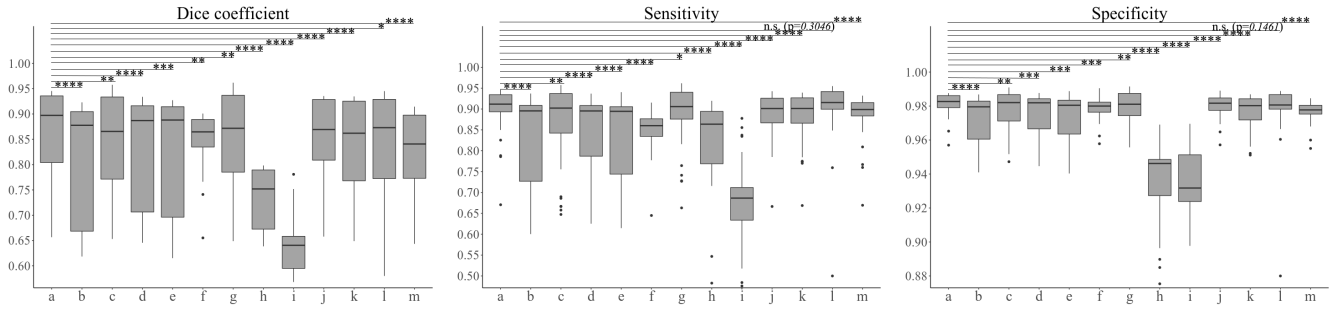


Fig. 8. Box plot depicting the importance of  $S_t$ NHMRF\_S algorithm for brain MR image segmentation in BrainWeb and IBSR databases over state-of-the-art segmentation methods. (a)  $S_t$ NHMRF\_S, (b) DSFCM\_N, (c) ICFFCM, (d) FRFCM, (e) KWFLICM, (f) BCFCM, (g) MICO, (h) ASeg, (i) mEM, (j) RFCM, (k) rRFCM, (l) SPM8, and (m) FSL.

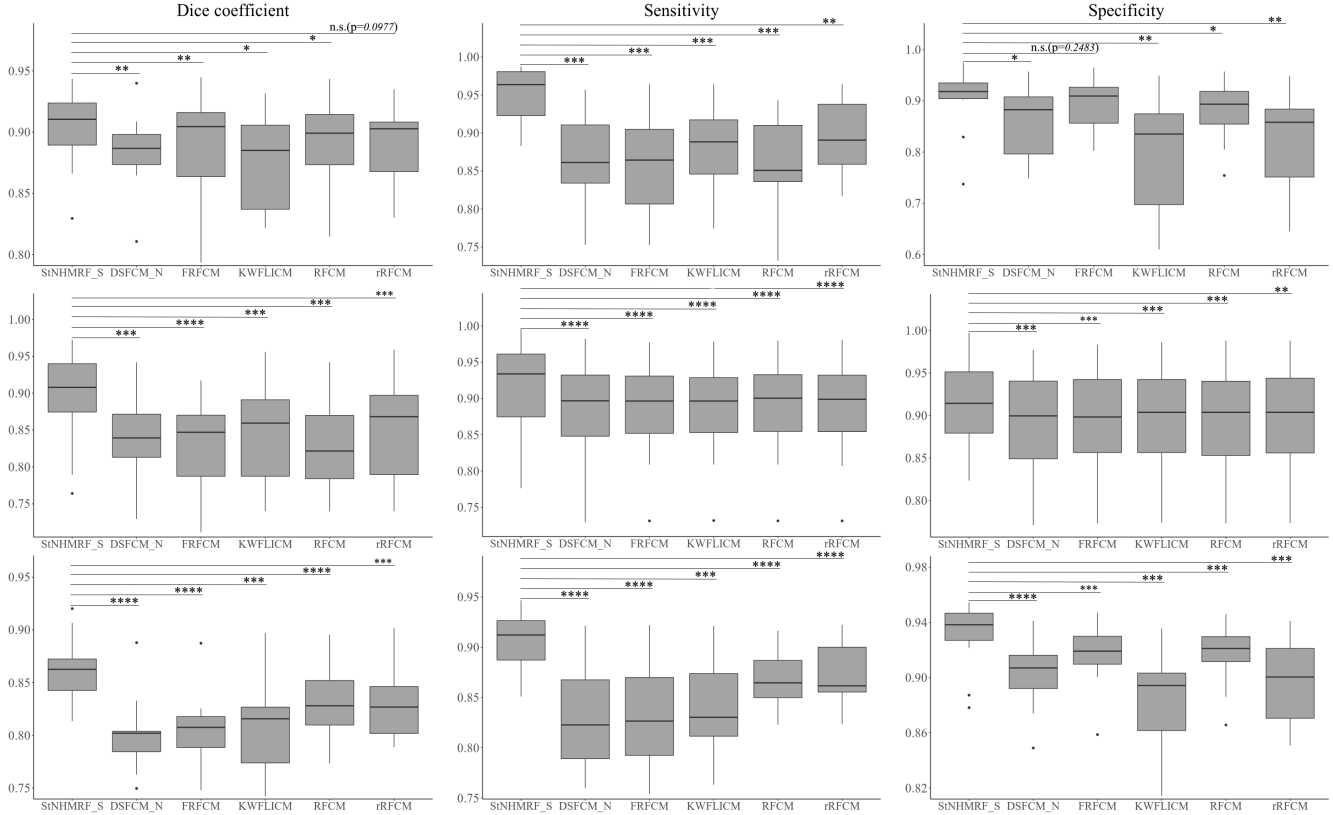


Fig. 9. Box plot depicting the importance of  $S_t$ NHMRF\_S algorithm over state-of-the-art segmentation methods: MIVIA database (top); Berkeley Segmentation Dataset (middle); and PASCAL VOC dataset (last).

only focuses on spatial distance of pixels with dissimilar class labels. In comparison, the proposed clique potential function integrates the information of image class parameters with the class label information of neighboring pixels and makes the contribution of spatial information compatible with that of pixel intensity information. Additionally, the new scaling parameter regulates the contribution of spatial information for each pixel based on its spatial location during the estimation of class labels and image class parameters.

2) *t*HMRF Model: The modification of *t*HMRF using proposed class label distribution attains significantly better segmented images compared to that obtained using the clique potential function of (9) for HEP-2 cell segmentation, natural image segmentation, and object segmentation. For brain MR

image segmentation, the proposed class label distribution provides significant performance improvement with respect to Dice coefficient and specificity, and while better but not significant improvement with respect to sensitivity. Compared to the clique potential function of (10), the proposed function of (26) achieves statistically significant improvement in segmentation performance for *t*HMRF algorithm, irrespective of quantitative indices and image databases used.

3) *G*HMRF Model: The modification of *G*HMRF using proposed class label distribution attains significantly better segmented images compared to that obtained using both clique potential functions of (9) and (10), irrespective of the image databases and segmentation evaluation indices used. Hence, all results reported in Fig. 3 establish the importance of using

new class label distribution, in terms of optimal segmentation performance. The segmented images using new class label distribution, along with the input image and corresponding ground-truth segmentation, are depicted in Figs. 4-7. The segmentation performance of algorithms using existing class label distributions of (9) and (10) are also presented in same figures. From all the quantitative and qualitative results, it is clear that the proposed class label distribution improves the segmentation performance of all HMRF based approaches, namely, GHMRF,  $t$ HMRF, and  $St$ NHMRF.

#### D. Importance of $St$ N Distribution

The incorporation of new class label distribution into existing HMRF model based algorithms improve the performance of the algorithms in terms of segmentation accuracy. In this subsection, a comparative analysis is presented among all the modified algorithms. As the  $St$ N distribution provides better modeling of brain MR image intensity distribution than the Student's- $t$  and Gaussian distributions, the  $St$ NHMRF\_S algorithm has achieved the optimal segmentation performance among other probabilistic model based algorithms.

From Fig. 3, it is observed that, for natural image segmentation, the  $St$ NHMRF\_S algorithm achieves significantly better segmentation results, compared to both  $t$ HMRF\_S and GHMRF\_S, irrespective of the segmentation evaluation indices. For brain MR image segmentation, HEp-2 cell segmentation, and object segmentation, the  $St$ NHMRF\_S algorithm achieves significantly better segmentation results, compared to both  $t$ HMRF\_S and GHMRF\_S, with respect to sensitivity and specificity; while better but not significant results are observed with respect to Dice coefficient. Hence, all results reported in Fig. 3 demonstrate the advantage of using  $St$ N distribution, in terms of optimal segmentation performance.

#### E. Performance of Different Algorithms

This section compares the segmentation performance of the  $St$ NHMRF\_S with that of several state-of-the-art segmentation methods. From all the results reported in Fig. 8 for brain MR image segmentation, it is observed that the  $St$ NHMRF\_S provides significantly better segmentation results compared to DSFCM\_N, ICFFCM, FRFCM, KWFLICM, BCFCM, MICO, ASeg, mEM, RFCM, rRFCM, and FSL, irrespective of quantitative evaluation indices. Compared to SPM8, the performance of  $St$ NHMRF\_S is significantly better with respect to Dice coefficient, while better but not significant with respect to both sensitivity and specificity. From all the results reported in Fig. 9 for HEp-2 cell segmentation, it is observed that the  $St$ NHMRF\_S provides significantly better segmentation results compared to DSFCM\_N, KWFLICM, and RFCM. The performance of  $St$ NHMRF\_S over FRFCM is significantly better with respect to both Dice coefficient and sensitivity and better but not significant with respect to specificity. Compared to rRFCM, the performance of  $St$ NHMRF\_S is also significantly better in terms of both sensitivity and specificity values; while better but not significant performance is observed with respect to Dice coefficient.

For both natural image segmentation and object segmentation, the proposed  $St$ NHMRF\_S algorithm achieves significantly better segmentation results compared to all state-of-the-art methods, namely, DSFCM\_N, FRFCM, KWFLICM, RFCM, and rRFCM, irrespective of quantitative evaluation indices. The segmented images by the proposed  $St$ NHMRF\_S algorithm and state-of-the-art methods, reported in Figs. 4-7, infer that the proposed  $St$ NHMRF\_S algorithm generates more promising results than do the state-of-the-art methods.

#### V. CONCLUSION

The problem of image segmentation requires special attention due to the presence of noise and several degrading artifacts, since they deteriorate the performance of segmentation methods. In this regard, the contribution of the paper is four-fold, namely, introducing a novel class label distribution, using the information of tissue class parameters, to achieve optimal labeling of image pixels as well as optimal estimate of tissue class parameters; introducing a novel scaling parameter, which facilitates HMRF model based segmentation algorithms to adaptively determine the contribution of spatial information during class label estimation of a specific pixel; modifying the probabilistic model based image segmentation algorithms, incorporating the novel class label distribution and scaling parameter; and illustrating the efficacy of the proposed algorithms, along with a comparison with related algorithms, qualitatively as well as quantitatively, on a set of synthetic and real brain MR images, HEp-2 cell IIF images, and natural images for class and object segmentation.

The proposed model has been shown to attain better image segmentation results, as it incorporates the information of class parameters into the class label distribution. Also, introduction of the new scaling parameter eradicates the ambiguity in classifying the edge pixels of the image classes. As  $St$ N distribution provides better modeling of the intensity distribution in an image compared to both Gaussian and Student's- $t$  distributions, incorporation of the novel class label distribution into  $St$ NHMRF framework has been able to outperform other probabilistic model based algorithms. Although the introduction of novel class label distribution and new scaling parameter has been illustrated for unsupervised image segmentation tasks, the concept can potentially be integrated with machine learning architecture in near future.

#### REFERENCES

- [1] W. M. Wells, III, W. E. L. Grimson, R. Kikins, and F. A. Jolezs, "Adaptive Segmentation of MRI Data," *IEEE Transactions on Medical Imaging*, vol. 15, no. 8, pp. 429–442, 1996.
- [2] Z. Liang, J. R. MacFall, and D. P. Harrington, "Parameter Estimation and Tissue Segmentation from Multispectral MR Images," *IEEE Transactions on Medical Imaging*, vol. 13, no. 3, pp. 441–449, 1994.
- [3] H. Greenspan, A. Ruf, and J. Goldberger, "Constrained Gaussian Mixture Model Framework for Automatic Segmentation of MR Brain Images," *IEEE Transactions on Medical Imaging*, vol. 25, no. 9, pp. 1233–1245, 2006.
- [4] T. M. Nguyen, Q. M. J. Wu, and S. Ahuja, "An Extension of the Standard Mixture Model for Image Segmentation," *IEEE Transactions on Neural Networks*, vol. 21, no. 8, pp. 1326–1338, 2010.
- [5] J. Liu and H. Zhang, "Image Segmentation Using a Local GMM in a Variational Framework," *Journal of Mathematical Imaging and Vision*, vol. 46, no. 2, pp. 161–176, 2013.

- [6] Y. Zhang, M. Brady, and S. Smith, "Segmentation of Brain MR Images Through a Hidden Markov Random Field Model and the Expectation-Maximization Algorithm," *IEEE Transactions on Medical Imaging*, vol. 20, no. 1, pp. 45–57, 2001.
- [7] K. Held, E. R. Kops, B. J. Krause, W. M. Wells, III, R. Kikinis, and H. W. Muller-Gartner, "Markov Random Field Segmentation of Brain MR Images," *IEEE Transactions on Medical Imaging*, vol. 16, no. 6, pp. 878–886, 1997.
- [8] A. Diplaros, N. Vlassis, and T. Gevers, "A Spatially Constrained Generative Model and an EM Algorithm for Image Segmentation," *IEEE Transactions on Neural Networks*, vol. 18, no. 3, pp. 798–808, 2007.
- [9] T. M. Nguyen and Q. M. J. Wu, "Fast and Robust Spatially Constrained Gaussian Mixture Model for Image Segmentation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 4, pp. 621–635, 2013.
- [10] A. Roche, D. Ribes, M. Bach-Cuadra, and G. Krüger, "On the Convergence of EM-Like Algorithms for Image Segmentation Using Markov Random Fields," *Medical Image Analysis*, vol. 15, pp. 830–839, 2011.
- [11] A. Banerjee and P. Maji, "Rough Sets and Stomped Normal Distribution for Simultaneous Segmentation and Bias Field Correction in Brain MR Images," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5764–5776, 2015.
- [12] —, "Spatially Constrained Student's  $t$ -Distribution Based Mixture Model for Robust Image Segmentation," *Journal of Mathematical Imaging and Vision*, vol. 60, no. 3, pp. 355–381, 2018.
- [13] J. Besag, "Spatial Interaction and the Statistical Analysis of Lattice Systems," *Journal of the Royal Statistical Society, Series B*, vol. 36, no. 2, pp. 192–326, 1974.
- [14] A. Banerjee and P. Maji, "Stomped- $t$ : A Novel Probability Distribution for Rough-Probabilistic Clustering," *Information Sciences*, vol. 421, pp. 104–125, 2017.
- [15] M. Woolrich and T. Behrens, "Variational Bayes Inference of Spatial Mixture Models for Segmentation," *IEEE Transactions on Medical Imaging*, vol. 25, no. 10, pp. 1380–1391, 2006.
- [16] J. Besag, "On the Statistical Analysis of Dirty Pictures," *Journal of the Royal Statistical Society, Series B*, vol. 48, no. 3, pp. 259–302, 1986.
- [17] Y. Zhang, X. Bai, R. Fan, and Z. Wang, "Deviation-Sparse Fuzzy C-Means With Neighbor Information Constraint," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 1, pp. 185–199, 2019.
- [18] X. Bai, Y. Zhang, H. Liu, and Y. Wang, "Intuitionistic Center-Free FCM Clustering for MR Brain Image Segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. PP, no. PP, pp. 1–10, 2019.
- [19] T. Lei, X. Jia, Y. Zhang, L. He, H. Meng, and A. K. Nandi, "Significantly Fast and Robust Fuzzy C-Means Clustering Algorithm Based on Morphological Reconstruction and Membership Filtering," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 5, pp. 3027–3041, 2018.
- [20] M. Gong, Y. Liang, J. Shi, W. Ma, and J. Ma, "Fuzzy C-Means Clustering With Local Information and Kernel Metric for Image Segmentation," *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 573–584, 2013.
- [21] M. N. Ahmed, S. M. Yamany, N. Mohamed, A. A. Farag, and T. Moriarty, "A Modified Fuzzy C-Means Algorithm for Bias Field Estimation and Segmentation of MRI Data," *IEEE Transactions on Medical Imaging*, vol. 21, no. 3, pp. 193–199, 2002.
- [22] C. Li, J. C. Gore, and C. Davatzikos, "Multiplicative Intrinsic Component Optimization (MICO) for MRI Bias Field Estimation and Tissue Segmentation," *Magnetic Resonance Imaging*, vol. 32, pp. 913–923, 2014.
- [23] R. Guillemaud and M. Brady, "Estimating the Bias Field of MR Images," *IEEE Transactions on Medical Imaging*, vol. 16, no. 3, pp. 238–251, 1997.
- [24] P. Maji and S. K. Pal, "Rough Set Based Generalized Fuzzy C-Means Algorithm and Quantitative Indices," *IEEE Transactions on System, Man, and Cybernetics, Part B: Cybernetics*, vol. 37, no. 6, pp. 1529–1540, 2007.
- [25] P. Maji and S. Paul, "Rough-Fuzzy Clustering for Grouping Functionally Similar Genes from Microarray Data," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 10, no. 2, pp. 286–299, 2013.
- [26] J. Ashburner and K. J. Friston, "Unified Segmentation," *NeuroImage*, vol. 26, no. 3, pp. 839–851, 2005.
- [27] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith, "FSL," *NeuroImage*, vol. 62, no. 2, pp. 782–790, 2012.
- [28] R. K.-S. Kwan, A. C. Evans, and G. B. Pike, "MRI Simulation-Based Evaluation of Image-Processing and Classification Methods," *IEEE Transactions on Medical Imaging*, vol. 18, no. 11, pp. 1085–1097, 1999.
- [29] D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sled, N. J. Kabani, C. J. Holmes, and A. C. Evans, "Design and Construction of a Realistic Digital Brain Phantom," *IEEE Transactions on Medical Imaging*, vol. 17, no. 3, pp. 463–468, 1998.
- [30] P. Foggia, G. Percannella, P. Soda, and M. Vento, "Benchmarking HEp-2 Cells Classification Methods," *IEEE Transactions on Medical Imaging*, vol. 32, no. 10, pp. 1878–1889, 2013.
- [31] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A Database of Human Segmented Natural Images and Its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics," in *Proceedings of the 8th IEEE International Conference on Computer Vision*, 2001, pp. 416–423.
- [32] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results," <http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html>.
- [33] S. M. Smith, "Fast Robust Automated Brain Extraction," *Human Brain Mapping*, vol. 17, no. 3, pp. 143–155, 2002.
- [34] A. Banerjee and P. Maji, "Rough Sets for Bias Field Correction in MR Images Using Contraharmonic Mean and Quantitative Index," *IEEE Transactions on Medical Imaging*, vol. 32, no. 11, pp. 2140–2151, 2013.
- [35] N. Otsu, "A Threshold Selection Method from Gray-Level Histogram," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.



**Abhirup Banerjee** received the BSc degree in Statistics from University of Calcutta, India in 2009 and Master degree in Statistics and Ph.D. degree in Computer Science both from Indian Statistical Institute, Kolkata, India, in 2011 and 2017, respectively. Currently, he is a postdoctoral researcher in the Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford and a stipendiary lecturer in the Mansfield College, Oxford, United Kingdom. His research interests include biomedical image analysis, machine learning, statistical pattern recognition, and so forth. He has published around 20 papers in international journals and conferences. He is also a reviewer of many international journals. Dr. Banerjee has received the ISCA Young Scientist Award from the Indian Science Congress Association in the year 2016-2017.



**Pradipta Maji** received the B.Sc. degree in physics, the M.Sc. degree in electronics science, and the Ph.D. degree in the area of computer science from Jadavpur University, Kolkata, India, in 1998, 2000, and 2005, respectively. Currently, he is a Professor with the Machine Intelligence Unit, Indian Statistical Institute, Kolkata, India. He has authored or coauthored more than 125 papers in international journals and conferences. His research interests include pattern recognition, machine learning, computational biology and bioinformatics, medical image processing, and so forth. Dr. Maji is a Fellow of the National Academy of Sciences, India. He has received the 2008 Microsoft Young Faculty Award from Microsoft Research Laboratory India Pvt., the 2009 Young Scientist Award from the National Academy of Sciences, India, the 2011 Young Scientist Award from the Indian National Science Academy, India, and the 2015 Young Faculty Research Fellowship from the Ministry of Electronics and Information Technology, Government of India. He has been selected as the 2009 Young Associate of the Indian Academy of Sciences, India.