

Fernandez-Steel Skew Normal Mixture Model Using Bayesian Approach for MRI-Based Brain Tumor Segmentation

Pravitasari, A. A. ^{*1,2}, Islamiyah, M. I. D.¹, Iriawan, N.¹,
Irhamah¹, Fithriasari, K.¹, Purnami, S. W.¹, and Ferriastuti, W.³

¹*Department of Statistics, Faculty of Science and Data Analytics,
Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia*

²*Department of Statistics, Faculty of Mathematics and Natural
Sciences, Universitas Padjadjaran, Bandung, Indonesia*

³*Department of Radiology, Faculty of Medicine, Universitas
Airlangga, Surabaya, Indonesia*

E-mail: nur_i@statistika.its.ac.id

** Corresponding author*

Received: 10 March 2019

Accepted: 27 March 2020

ABSTRACT

Detection of a brain tumor in Magnetic Resonance Imaging (MRI) is always challenging due to the gray level comparison of tumor and normal tissue. Model-based clustering with a Finite Mixture Model (FMM) is widely used to segment the tumor as the Region of Interest (ROI). The Gaussian Mixture Model (GMM) is becoming abandoned because, in reality, the symmetric distribution approach is less able to explain the MRI data pattern. In addition, the use of a symmetric distribution cannot compete for the model parsimonious of an asymmetric distribution to exhibit the long and heavy tail pattern of the data. On this kind of data, more Gaussian mixture components are needed in the GMM.

This study, therefore, develops a mixture model with asymmetric distribution, called Fernandez-Steel Skew Normal (FSSN). It is one of the Neo-Normal distributions that can be skewed adaptively but remains stable in its mode of distribution. Bayesian coupled with the Markov chain Monte Carlo (MCMC) approach is employed for estimating FSSN distribution parameters numerically. Silhouette Index (SI) coefficient is performed to validate the result of the segmentation. The results indicate that the FSSN mixture model (FSSN-MM) has a better performance at representing the data pattern of a brain tumor MRI. This is indicated by the higher *SI* coefficient of the FSSN-MM than GMM. In addition, the FSSN-MM is more parsimonious, since it has the smallest number of clusters. Moreover, FSSN-MM is able to detect the brain tumor more precisely than the original GMM approach.

Keywords: Bayesian, brain tumor, image segmentation, Fernandez-steel skew normal and mixture model.

1. Introduction

The brain is a vital organ which has the ability to control every human activity. One disease that attacks the brain is a tumor. It occurs due to the proliferation and rapid growth of abnormal cells in the central nervous system (CNS) and brain wrapping membranes (meningeal membranes) (ABTA, 2018). According to World Health Organization (WHO), there were 5,323 cases of brain and nervous system tumor in Indonesia during 2018. Brain tumors amount to the 15th largest cause of death compared to all types of cancer, both in men and women. The total mortality (number of deaths) in 2018 due to brain tumors is 4,229 cases (Globocan, 2018). This statistical information makes an important excuse to develop this study.

An MRI is a form of digital imaging technology which is often used by health experts to detect brain tumors. Examination of MRI images requires precision and accuracy, and there are difficulties in processing these images. One such difficulty is separating one object that's considered medically more important, namely it the Region of Interest or ROI, than another object (Non-ROI) (Angulakshmi and LakshmiPriya, 2018). This process is called MRI image segmentation that separating the image of the brain tumor as ROI from other background images as Non-ROI.

Many methods have been developed for image segmentation; one of them is Model-Based Clustering which based on a probability of the data and constructed as the FMM. GMM is one of the FMM which are widely used for image segmentation. The disadvantage of using GMM relates to its parsimonious issues. The Gaussian distribution that constructs the GMM has a short-tailed characteristic. When the data pattern has a longer tail, it would be approached with more components of the Gaussian distribution to form a mixture model.

An alternative to this problem is using the Neo-normal distribution as a replacement for the Gaussian distribution. This distribution is an adaptive distribution which is more flexible to capture both the symmetrical and the skew data pattern (Pravitasari et al., 2019a). One of the Neo-Normal family is the FSSN distribution developed by Fernandez and Steel (1998). FSSN have succeeded to improve the previous Neo-Normal distribution, i.e. The Exponential Power (EP) distribution (Box and Tiao, 1992) and Azzalini Skew Normal (Azzalini, 1985). EP distribution very smart to capture all of leptokurtic, mesokurtic, and platikurtic adaptively to data, in contrast it blinds to skewness of data. While Azalini Skew Normal is able to accommodate the skewness but its not stable in its location. Another Neo-normal distribution developed by Iriawan (2000) namely Modified Stable Burr (MSBurr), when MS-

Burr approximated the normal density called MSNBurr, while approximated Student's t density called MSTBurr. FSSN distribution is a distribution that forms a normal or Gaussian distribution or Student's t that can be skewed adaptively but still stable in its mode (Castillo et al., 2011). Therefore, this study will employ the FSSN distribution in composing a mixture model; named as the FSSN-MM, and it's expected to produce better segmentation and more parsimonious than the original GMM approach. Last but not least, the optimization to estimate model parameters is used by employing Bayesian coupled with the MCMC approach, since the classical approach did not provide the closed form solutions.

2. Fernandez-Steel Skew Normal Mixture Model (FSSN-MM)

In this research, the method used for image segmentation of MRI brain tumor is FSSN-MM. Suppose that ε is the residual model that spread in the domain $-\infty < \varepsilon < \infty$ and is normally distributed, $\varepsilon(0, \sigma^2)$. FSSN distribution for a variable ε according to Fernandez and Steel (1998), is defined by equation (1).

$$f(\varepsilon|\gamma) = \frac{2}{\gamma + \frac{1}{\gamma}} \left\{ f\left(\frac{\varepsilon}{\gamma}\right) I_{[0, \infty)}(\varepsilon) + f(\gamma\varepsilon) I_{(-\infty, 0)}(\varepsilon) \right\}, \gamma = (0, \infty) \quad (1)$$

where γ is the skewness or transformation parameter. This distribution will lose its symmetry when $\gamma \neq 1$, skew to the left when $\gamma < 1$ and skew to the right when $\gamma > 1$. Both patterns indicate that the mode of distribution is still stable. Based on the equation (1), the basic distribution that forms is a normal or Gaussian distribution, where the Gaussian density of ε is shown in equation (2).

$$f(\varepsilon|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\varepsilon - \mu}{\sigma}\right)^2\right] \quad (2)$$

where μ is the location parameter and σ is the scale parameter, $-\infty < \varepsilon < \infty$, $-\infty < \mu < \infty$ and $\sigma > 0$. In this study, the data used are MRI images of brain tumors that were provided by RSUD Dr. Soetomo Surabaya, Indonesia. Suppose y_i , $i = 1, 2, \dots, n$ are the grayscale intensities in pixels of the MRI image. If y_1, y_2, \dots, y_n follows the FSSN distribution, then the densities should be given by

$$f(y|\mu, \sigma, \gamma) = \frac{2}{\gamma + \frac{1}{\gamma}} \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} \left\{ \frac{1}{\gamma^2} I_{[\mu, 255)}(y) + \gamma^2 I_{(0, \mu)}(y) \right\} \right) \quad (3)$$

where $0 \leq y \leq 255$, $0 \leq \mu \leq 255$, $\sigma > 0$, and $\gamma > 0$.

The FSSN distribution in equation (3) will be used to construct the mixture model, called FSSN mixture model, for which the densities is provided by equation (4).

$$\begin{aligned}
 f(y|\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\gamma}, \mathbf{w}) &= \sum_{j=1}^K w_j f_j(y|\mu_j, \sigma_j, \gamma_j) \\
 &= \sum_{j=1}^K \frac{2w_j}{\gamma_j + \frac{1}{\gamma_j}} \left(\frac{e^{-\frac{1}{2}\left(\frac{y-\mu_j}{\sigma_j}\right)^2}}{\sigma_j \sqrt{2\pi}} \left\{ \frac{1}{\gamma_j^2} I_{[\mu, 255)}(y) + \gamma_j^2 I_{(0, \mu)}(y) \right\} \right), \tag{4}
 \end{aligned}$$

where w_j is the proportion parameter of the mixture component that satisfies $0 \leq w_j \leq 1, j = 1, 2, \dots, K$, and $\sum_{j=1}^K w_j = 1$.

2.1 Bayesian MCMC for FSSN-MM

The Markov chain Monte Carlo (MCMC) is one of the optimization methods which is good for estimating the parameter of FMM (Pravitasari et al., 2019b). In order to meet the requirement of MCMC, we need to design the prior distribution for each parameter as follows:

1. Prior distribution for $\mu_j, j = 1, 2, \dots, K$, according to Gelman et al. (2014), is Gaussian(η_j, φ_j). While prior for $\sigma_j, j = 1, 2, \dots, K$ is Invers Gamma (α_j, β_j).
2. Prior distribution for $\gamma_j, j = 1, 2, \dots, K$, according to Fernandez and Steel (1998), is Gamma(a_j, b_j).
3. Gelman et al. (2014) describes that the prior distribution for parameters of mixture proportions $w_j, j = 1, 2, \dots, K$ is the Dirichlet($\delta_1, \delta_2, \dots, \delta_K$) distribution.

While using the mixture model, each pixel $y_i, i = 1, 2, \dots, n$ has a latent variable $\mathbf{z}_i, i = 1, 2, \dots, n$, where $\mathbf{z}_i = (z_{ij})_{j=1}^K = (z_{i1}, z_{i2}, \dots, z_{iK})$, which is the allocation of each observation in each sub-population of the mixture model. A latent z_{ij} will be worth one, i.e. $z_{ij} = 1$, if the i^{th} pixel is allocated in the j^{th} sub-population. Otherwise $z_{ij} = 0$. So the likelihood function of the

FSSN-MM with a latent variable \mathbf{z}_i is defined in equation (5).

$$f(\mathbf{y}, \mathbf{z} | \mu_j, \sigma_j, \gamma_j, w_j) = \prod_{i=1}^n \prod_{j=1}^K (w_j f_j(y | \mu_j, \sigma_j, \gamma_j))^{z_{ij}}. \quad (5)$$

Full conditional posterior distribution is used to generate parameter values that will be estimated by the Bayesian MCMC. In this method, the full conditional posterior of each parameter is derived from the joint posterior calculated by multiplying the likelihood and the prior distribution of parameters which are estimated. The full conditional posterior of each parameter is explained below (all equations are written in logarithmic form for simplicity).

1. Equation (6) shows the full conditional posterior for location parameter $\mu_j, j = 1, 2, \dots, K$

$$\log f(\mu_j | \sigma_j, \gamma_j, w_j, \mathbf{z}, \mathbf{y}) = \text{constant} - \frac{1}{2} \left(\frac{\mu_j - \eta_j}{\varphi_j} \right)^2 - \frac{1}{2} \sum_{i=1}^n \left(\frac{y_i - \mu_j}{\sigma_j} \right)^2 \quad (6)$$

where all parts of the equation that do not contain the parameter μ_j are compiled as constant.

2. Equation (7) shows the full conditional posterior for parameter $\sigma_j, j = 1, 2, \dots, K$

$$\log f(\sigma_j | \mu_j, \gamma_j, w_j, \mathbf{z}, \mathbf{y}) = \text{constant} - (\alpha_j + 1) \log(\sigma_j) - \frac{\beta_j}{\sigma_j} - n \log \left(\sigma_j \gamma_j \sqrt{2\pi} + \frac{\sigma_j \sqrt{2\pi}}{\gamma_j} \right) - \sum_{i=1}^n \frac{\left(\frac{y_i - \mu_j}{\sigma_j} \right)^2}{2}, \quad (7)$$

where all parts of the equation that do not contain the parameter σ_j are collected together as a constant.

3. Equation (8) shows the full conditional posterior for parameter $\gamma_j, j = 1, 2, \dots, K$

$$\log f(\gamma_j | \mu_j, \sigma_j, w_j, \mathbf{z}, \mathbf{y}) = \text{constant} + (a_j - 1) \log(\gamma_j) - (b_j \gamma_j) + \sum_{i=1}^n \log \left(\frac{1}{\gamma_j^2} I_{[\mu, 255]}(y_i) + \gamma_j^2 I_{(0, \mu)}(y_i) \right) - n \log \left(\sigma_j \gamma_j \sqrt{2\pi} + \frac{\sigma_j \sqrt{2\pi}}{\gamma_j} \right), \quad (8)$$

where all parts of the equation that do not contain the parameter are summed as a constant.

4. Equation (9) shows posterior distribution for parameter $w_j, j = 1, 2, \dots, K$

$$f(w_j | \mathbf{y}, \mathbf{z}, n_j, \alpha_j, \phi_j) \propto \prod_{j=1}^K w_j^{\sum_{i=1}^n z_{ij}} = \prod_{j=1}^K w_j^{\left(\sum_{i=1}^n z_{ij} + 1\right) - 1}. \quad (9)$$

This is a Dirichlet $\left(1 + \sum_{i=1}^n z_{i1}, \dots, 1 + \sum_{i=1}^n z_{iK}\right)$ distribution.

5. The latent variable $z_{ij}, i = 1, 2, \dots, n$ and $j = 1, 2, \dots, K$, only has two possibilities value, that is 0 or 1. For certain i^{th} data, the j^{th} cluster will be set to 1 based on the most likely of i^{th} data to each Bernoulli $\left(\frac{f(y_i | \mu_j, \sigma_j, \gamma_j) w_j}{f(y_i)}\right)$, $j = 1, 2, \dots, K$. Therefore $\mathbf{z}_i = (z_{i1}, z_{i2}, \dots, z_{iK})$ will follow the Multinomial $(1, \omega_{i1}, \dots, \omega_{iK})$ distribution where

$$\omega_{ij} = \frac{f(y_i | \mu_j, \sigma_j, \gamma_j) w_j}{f(y_i)}. \quad (10)$$

Algorithm 1 states the steps of the Gibbs sampling framework as the process of Bayesian MCMC to optimize the FSSN-MM.

Algorithm 1. Gibbs Sampling Algorithm for FSSN-MM

1. Set the initial value for the parameters $\mu_1^0, \dots, \mu_K^0, \sigma_1^0, \dots, \sigma_K^0, \gamma_1^0, \dots, \gamma_K^0, w_1^0, \dots, w_K^0, z_1^0, \dots, z_n^0$
2. Update each variable in turn at the t^{th} iteration, $t = t + 1$
 - a Update value $\mu_j, j = 1, 2, \dots, K$ by generating μ_j^t according to equation (6)
 - b Update value $\sigma_j, j = 1, 2, \dots, K$ by generating σ_j^t according to equation (7)
 - c Update value $\gamma_j, j = 1, 2, \dots, K$ by generating γ_j^t according to equation (8)
 - d Update value $w_j, j = 1, 2, \dots, K$ by generating w_j^t that follow the Dirichlet $\left(1 + \sum_{i=1}^n z_{i1}, \dots, 1 + \sum_{i=1}^n z_{iK}\right)$ distribution.
 - e Update value z_i by generating z_i^t that follows the Multinomial $(1, \omega_{i1}, \omega_{i2}, \dots, \omega_{iK})$ distribution, where

$$\omega_{ij} = \frac{f(y_i | \mu_j, \sigma_j, \gamma_j) w_j}{f(y_i)}$$

for $i = 1, 2, \dots, n, j = 1, 2, \dots, K$.

3. Repeat step 2 for $t = 1, 2, \dots, T$ (T is number of samples generated).

2.2 Cluster Validation

Cluster validation is required to determine the optimum number of clusters. An index to validate a number of clusters is the Silhouette Index (SI). The SI for every i^{th} data at j^{th} cluster is formulated in equation (11) (Thinsungnoena *et al.*, 2015):

$$SI_i^j = \frac{b_i^j - a_i^j}{\max(a_i^j, b_i^j)}, \quad (11)$$

where a_i^j is the degree of similarity of a data is to its cluster, while b_i^j is the degree of similarity of a data to the other clusters.

The SI for each j^{th} cluster is the average value of all the data included in the j^{th} cluster. It can be formulated by equation (12). While the overall SI of the image is the average of all SI clusters formulated by equation (13).

$$SI_j = \frac{1}{m_j} \sum_{i=1}^{m_j} SI_i^j \quad (12)$$

$$SI = \frac{1}{K} \sum_{j=1}^K SI_j \quad (13)$$

where m_j is the total number of each data which is member of the j^{th} cluster. SI has a range of values $[-1, 1]$, the closer the SI to 1, the more precise the data in the cluster.

3. Results and Discussions

In this research, an MRI image used is determined by the medical recommendations. The chosen image sequence is ax T1 memp+C. It is one of the axial layers of the brain scan which adds contrast to enhance the tumor area. The preprocessing is conducted to improve the image quality and eliminate skull bones for easier segmentation. Figure 1 (a) and Figure 1 (b) show the MRI image, before and after preprocessing.

A histogram of a grayscale image can describe the distribution of its pixel intensity. Figure 2 (a) and Figure 2 (b) show the histogram of the MRI image before and after preprocessing. The histogram after preprocessing, has an asymmetrical shape and indicates skewness. Furthermore, histogram in Figure 2 (b) appears to have three modes, even though the modes of ROI pixels somehow look smaller and less visible. Thus, applying the mixture model framework

is considered appropriate in this case, as the histogram shows multimodality. In this study, the image is segmented using the FSSN-MM and compared with the GMM. The GMM used in this study are adapted from Sianipar (2017).

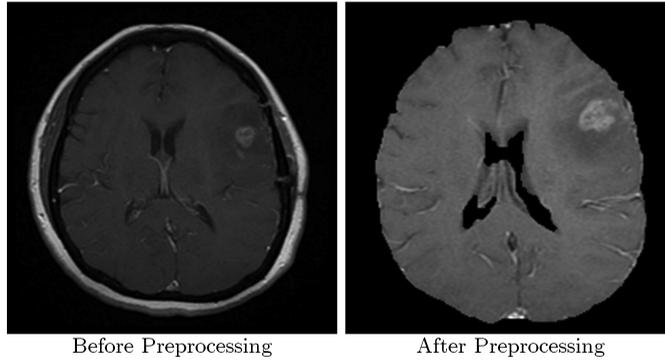


Figure 1: Image of (a) Original image of MRI sequence ax T1 memp+C, and (b) The sequence after preprocessing

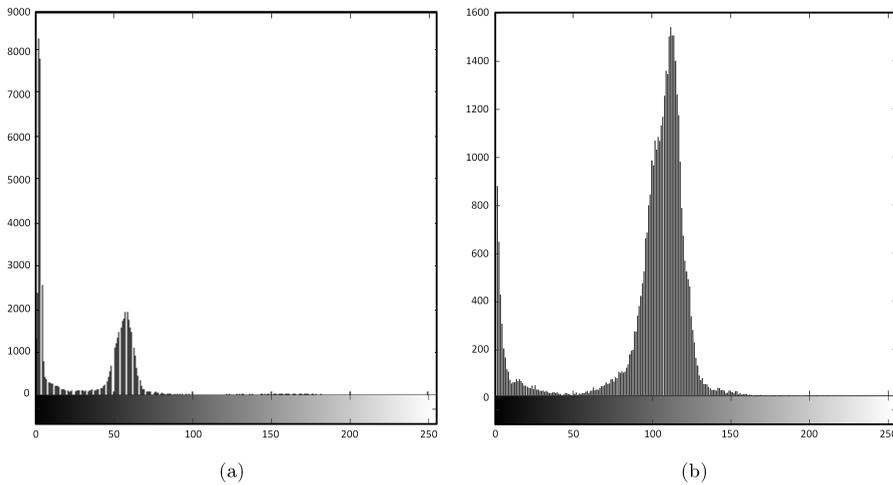
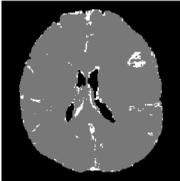
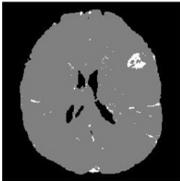
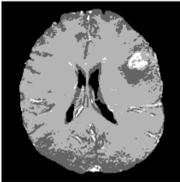
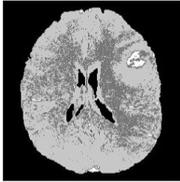
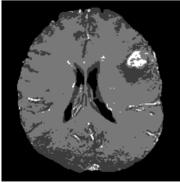
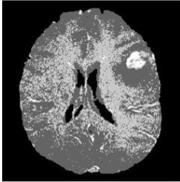
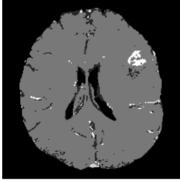
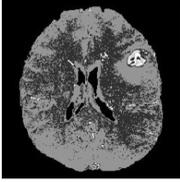
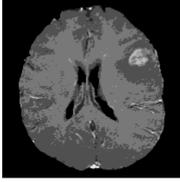
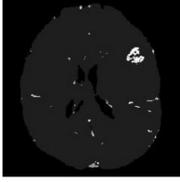


Figure 2: Histogram of (a) Original image of MRI sequence ax T1 memp+C, and (b) The histogram after preprocessing

Table 1 shows the segmentation results for both the GMM and FSSN. The method was used for $K = 2, 3, \dots, 7$ clusters. For $K = 2$ clusters, the segmen-

tation for both the GMM and the FSSN-MM is unable to capture the tumor area. This 2-components cluster only separates the brain from the background. Therefore, it will not be included in the analysis. Figure 3 shows the SI for both the GMM and the FSSN-MM. The bigger the SI value, the more optimum the number of clusters.

Table 1: Result of segmentation between GMM and FSSN-MM

K	GMM	FSSN-MM
3		
4		
5		
6		
7		

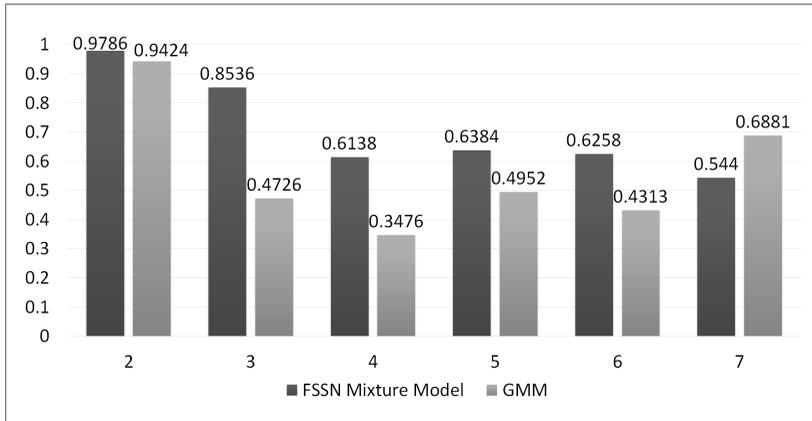


Figure 3: Silhouette Index of FSSN-MM and GMM

As in Figure 3, it can be determined that FSSN-MM reach the optimum in two number of clusters, while GMM in seven number of clusters. GMM gives a higher cluster number than the FSSN-MM, this indicates that the FSSN is more parsimonious than GMM due to its capability to explain the data pattern better despite having only three clusters. Their comparison of the estimated parameters is given in Table 2.

Table 2: Comparison of parameter model of GMM and FSSN-MM

K index	FSSN-MM				GMM		
	w	μ	σ	γ	w	μ	σ
1	0.440	1.679	23.579	4.468	0.522	1.669	0.748
2	0.551	108.31	63.879	12.103	0.030	89.949	16.86
3	0.009	148.04	76.028	14.405	0.230	105.453	9.959
4					0.197	114.500	7.341
5					0.014	127.675	14.620
6					0.004	147.374	11.599
7					0.002	171.629	34.498

In Table 2, the ROI in FSSN-MM lies in the third cluster, while in GMM is located in the seventh cluster. ROI in both methods is represented by the largest grayscale value, that is $\mu = 148.04$ for FSSN-MM and $\mu = 171,629$ for GMM. From the parameters w and σ it can be determined that the number of ROI pixels in the FSSN-MM is greater and has larger range of grayscale than the GMM. Especially for FSSN-MM, the γ parameter indicates that the distribution of each cluster is right skew.

Figure 4 demonstrates the visualization of the image before and after segmentation using the GMM and FSSN-MM. The segmentation results for both methods are chosen from its optimum number of clusters. This figure also shows how image segmentation using FSSN-MM clarifies the ROI area which is a significant contribution of this research. This result can be used as a suggestion for paramedics to detect the location of brain tumors more precisely.

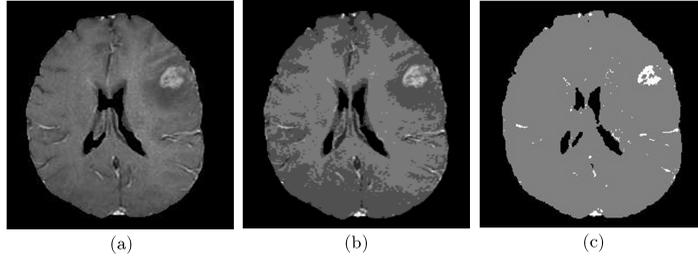


Figure 4: The segmentation result in the optimum number of cluster; (a) Without segmentation, (b) GMM with seven clusters, and (c) FSSN-MM with three clusters

4. Conclusions

Based on the results and discussion, the FSSN-MM gives better results of segmentation than the GMM. This is indicated by the *SI* coefficient of FSSN-MM which is higher than GMM. Another result also shows that the FSSN-MM is more parsimonious than GMM, as it has a smaller number of clusters, the FSSN-MM has been already able to capture the pattern of the original image. For further research, we suggest adding a more spatial approach to image segmentation. This is due to the FSSN-MM segmentation still containing a lot of noise which should be eliminated for a better result.

Acknowledgement

The Authors are grateful to the Directorate for Research and Community Service (DRPM), Ministry of Research, Technology and Higher Education Indonesia which support this research under PSN research grant no. 965/PKS/ITS/2018.

References

- ABTA (2018). About brain tumors: A primer for patients and caregivers. Available at https://www.abta.org/wp-content/uploads/2020/06/AboutBrain-Tumors_2020_web_en.pdf.
- Angulakshmi, M. and LakshmiPriya, G. G. (2018). Brain tumour segmentation from MRI using superpixels based spectral clustering. *Journal of King Saud University-Computer and Information Sciences*.(in press). <https://doi.org/10.1016/j.jksuci.2018.01.009>.
- Azzalini, A. (1985). A class of distribution which includes the normal ones. *Scandinavian Journal of Statistics*, 12(2):171–178.
- Box, G. and Tiao, G. C. (1992). *Bayesian Inference in Statistical Analysis*. New York: John Wiley & Sons.
- Castillo, N. O., Gomez, H. W., Leiva, V., and Sanhueza, A. (2011). On the Fernandez-steel distribution: Inference and application. *Computational Statistics and Data Analysis*, 55(11):2951–2961.
- Fernandez, C. and Steel, M. F. J. (1998). On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93(441):359–371.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian Data Analysis*. Boca Raton: CRC Press.
- Globocan (2018). Indonesia cancer fact sheets. <http://gco.iarc.fr/today/data/factsheets/populations/360-indonesia-factsheets.pdf>.
- Iriawan, N. (2000). *Computationally Intensive Approaches to Inference in Neo-Normal Linier Models*. PhD thesis, Curtin University of Technology, Australia.
- Pravitasari, A. A., Iriawan, N., Safa, M., Fithriasari, K., Irhamah, Purnami, S. W., and Ferriastuti, W. (2019a). MRI-based brain tumor segmentation using modified stable student's t from burr mixture model with bayesian approach. *Malaysian Journal of Mathematical Science*, 13(3):297–310.
- Pravitasari, A. A., Nirmalasari, N. I., Iriawan, N., Fithriasari, K., Irhamah, Purnami, S. W., and Ferriastuti, W. (2019b). Bayesian spatially constrained Fernandez-steel skew normal mixture model for MRI-based brain tumor segmentation. *AIP Conference Proceedings*, 2184:020086(1–9). <https://doi.org/10.1063/1.5139818>.

- Sianipar, W. H. G. (2017). *Comparison Between EM-GMM (Expectation maximization-Gaussian Mixture Model) for Brain MRI Segmentation in RSUD Soetomo to Located Brain Tumour Area*. Undergraduate thesis, Institut Teknologi Sepuluh Nopember, Surabaya.
- Thinsungnoena, T., Kaoungkub, N., Durongdumronchaib, P., Kerdprasopb, K., and Kerdprasopb, N. (2015). The clustering validity with silhouette and sum of squared errors. *Proceedings of the 3rd International Conference on Industrial Application Engineering*, 2015:44–51.