Clustering Techniques for Segmentation of Soft Tissue Sarcoma in MR Images

Koushik Chakraborty, Department of Electronics and Communication Engineering, Jayoti Vidyapeeth Women's University, Jaipur, India. E- mail: Koushik215@gmail.com

Tapas Si, Department of Computer Science and Engineering, Bankura Unnayani Institute of Engineering, Bankura, West Bengal, India. E-mail: c2.tapas@gmail.com

Arunava De, Department of Electronics and Communication Engineering, K.L.E.F., Guntur, Andhra Pradesh, India.

E-mail: arunavade@yahoo.com

Sudhir K. Sharma, Department of Electronics and Communication Engineering, Jaipur National University, Jaipur, India.

E-mail: sudhir.732000@gmail.com

Abstract --- This paper presents soft tissue sarcoma segmentation in MR images using hard-clustering techniques with Swarm Intelligence (SI). Swarm Intelligence based clustering techniques are popular and widely used in segmentation of brain MRI. In this work, the noise in MRI is reduced using median filter and intensity in homogeneities is corrected in the preprocessing steps. In the subsequent steps the MR images are segmented using hard-clustering technique using Swarm Intelligence algorithms. Qualitative and quantitative performances are analyzed by applying clustering for segmentation of MR images using five SI algorithms such as Particle Swarm Optimizer, Artificial Bee Colony, Fire-Fly algorithm, Glowworm Swarm Optimizer with Generalized Opposition Based Learning, Fireworks algorithm with Adaptive Transfer Function.

Keywords--- Soft Tissue Sarcoma, MRI, Swarm Intelligence, Clustering Techniques.

I. Introduction

Soft tissue sarcomas (STS) are malignant tumors that originate in tissues like muscle, fat, nerves, fibrous tissues or blood vessels. Segmentation of soft tissue sarcomas in MR images is a very important medical image analysis task for diagnostic, treatment and follow-up the disease.

There are some interesting contributions to the segmentation of sarcoma.J.Juntu et al.[3] extracted texture features from T1-weighted MRI and classification of benign and malignant tumors were done using applied machine learning methods. H. Farhidzadeh et al. [4] proposed Otsu method for segmentation of soft tissue sarcoma. H. Farhidzadeh et al. [5] proposed Fuzzy c-means based segmentation method for soft tissue sarcoma. H. Farhidzadeh et al. [6] predicted necrotic as well as metastatic soft tissue sarcomas using texture feature analysis.

M.Vallieres et al. [7] proposed a method of predicting lung metastases using MRI texture features and FDG-PET for soft tissue sarcomas. J.O. Glass and W. E. Reddick [8] proposed a segmentation method for osteo-sarcoma using Kohonen self-organizing map (SOM) and multilayer back propagation neural network. In soft tissue sarcoma, volumetric quantification of tumor necrosis was done by W. L. Monsky et al. [9] using an iterated watershed method.

Nature-inspired strategies e.g. swarm intelligence (SI) and artificial neural network **are** now being used in decision making systems for solving the medical image analysis problems [11]. The objective of this paper is to make a comprehensive study of the performance of five SI based clustering techniques in the segmentation of soft tissue sarcomas in MR images. The five SI algorithms are Particle Swarm Optimizer (PSO), Artificial Bee Colony (ABC), Fire-fly algorithm (FFA), Glowworm Swarm Optimizer with generalized opposition based learning (GOBL-GSO) and Fireworks algorithm with adaptive transfer function (FWA-ATF).

II. Materials and Methods

2.1MRI Dataset

The MRI datasets [20-22] of two patients are used in this study. The MR images are collected from "The Cancer Imaging Archive" [22]. Each dataset in this study contains ten MR images.

Received: 20 September 2018/Accepted: 15 October 2018

This dataset contains histologically proven soft-tissue sarcomas (STSs), FDG-PET/C, T1-weighted, T2-weighted with fat-suppression MRI data. In this study, T2-weighted with fat-suppression MRI data of two patients are used.

2.2 Noise Reduction

Segmentation of soft tissue sarcomas faces difficulties on account of presence of noise in the MR images. Median filter with neighborhood size 3x3[10] is used to remove the noise. Median filter is used as a de-noising method for noise reduction in the MR image and it preserves the edges of the MR image [10].

2.3Intensity In homogeneity Correction

The intensity non-uniformity or the bias field usually refers to the intensity variations of the same tissue over the image due to the radio-frequency no uniformity, static field in homogeneity etc .in the imaging instrumentation, or the patient movement. A 10x10 Max filter based inhomogeneity correction method [16] is used to correct intensity in-homogeneities in the denoised images.

2.4Segmentation using SI based Clustering Algorithm

2.4.1 SI based Clustering Algorithm

SI based clustering technique is a popular hybrid clustering techniques used for segmentation of MR images. The SI algorithms such as PSO [13], ABC[14], FFA[15], GOBL-GSO[16], FWA-ATF[17]are widely used in clustering techniques and they perform better than well-known K-means algorithm[12]. All these algorithms use three objective functions for better clustering and they are discussed in the next section.

2.4.2 Objective Functions

In this present work, in order to get better clustering solutions, all the existing SI based clustering algorithms used three objective functions. The first objective function f_1 has to be minimized. f_1 is the maximum intra-cluster distance. It is defined as follows:

$$f_1 = d_{max}(X,\vec{m}) = \max_{\forall k \in K}\left\{\sum_{\forall \vec{x} \in C_k} \frac{d(\vec{x_i},\vec{m_k})}{|C_k|}\right\}$$

Where $|C_k|$ is the number of pixels belonging to the cluster C_k. f_2 is the minimum inter-class separation. It has to be maximized and it is defined as follows:

$$f_2 = \min_{\substack{i \neq j \\ i,j \in k}} d(\vec{m_i},\vec{m_j})$$

The third objective function f_3 is the quantization error which has to be minimized.

$$f_3 = \frac{1}{K}\sum_{k=1}^{K}\sum_{\forall \vec{x} \in C_k} \frac{d(\vec{m_i},\vec{m_j})}{|C_k|}$$

All of the above objective functions are combined into one objective function (F) which has to be minimized and it is given by:

$$F = w_1.f_1 + w_2.(x_{max} - f_2) + w_3.f_3$$

Where $w_i \in (0,1)$ is the weight value and $\sum_{i=1}^{3} w_i = 1$. f_2 is converted into a minimization problem by deducting it from the highest pixel intensity x_{max} of the MR images.

2.5 Performance Evaluation

Davis–Bould in (DB) *index* [18] has been used to evaluate the segmentation performance quantitatively. DB-*Index* is widely used cluster validity index.

Ratio of the sum of intra-cluster scatter to inter-cluster separation is DB-Index. The i^{th} intra-cluster scatter is defined as follows:

$$S_{i,q} = \left[\frac{1}{N_i}\sum_{\forall \vec{x} \in C_k} |\vec{x} - \vec{m_i}|^q\right]^{\frac{1}{q}}$$

Received: 20 September 2018/Accepted: 15 October 2018

Inter-cluster distance between i[th] and j[th] cluster is as follows:

$$D_{ij} = \left[\sum_{p=1}^{d} |m_{i,p} - m_{j,p}|^t\right]^{\frac{1}{t}}$$

Where $\overrightarrow{m_i}$is the center of i^{th} cluster, q, t >=1, q, t are integer, $N_i = |C_i|$is the number of data points in the i^{th}clusterC_i. q and t are set to 2 in this present work. $R_{i,qt}$ is calculated as follows:

$$R_{i,qt} = \max_{\substack{j \in K \\ j \neq i}} \left\{ \frac{S_{i,q} + S_{j,q}}{D_{ij}} \right\}$$

Finally, DB-Index is measured as follows:

$$DB(K) = \frac{1}{K} \sum_{i=1}^{K} R_{i,qt}$$

III. Results and Discussion

The classical K-means, PSO, ABC, FFA, GOBL-GSO and FWA-ATF based clustering techniques are applied on ten MR slices of each patient. The clustering techniques are applied to segment the MR images into three classes: (i) enhanced or active regions, (ii) non-enhanced or non-active sub regions and (iii) region includes other tissues and background.

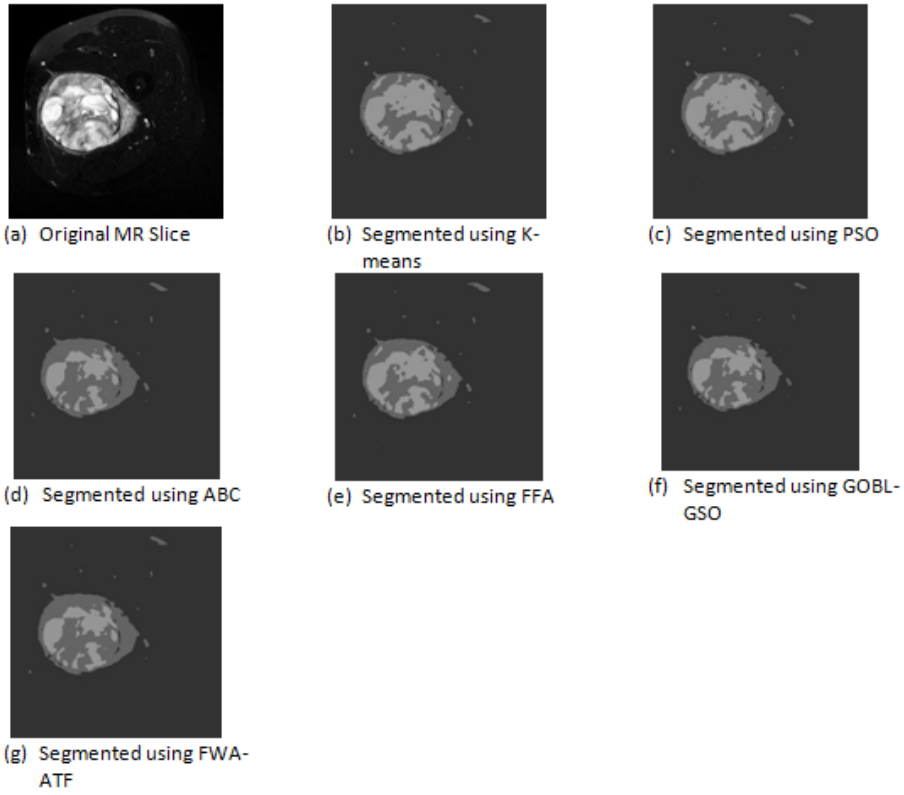

Figure 1: Original MR Slice and its Segmented Image Using Different Methods

The qualitative results are given in Fig.1 (b)-(g) in terms of segmented image of one original image (Fig.1 (a)). In the segmented images, highest intensity indicates enhanced or active region of the sarcomas whereas medium level intensity indicates non-enhanced or non-active regions of the sarcomas. From Fig.1 (b)-(g), it is visually observed that ABC, GOBL-GSO and FWA-ATF based clustering techniques perform better segmentation than other

Received: 20 September 2018/Accepted: 15 October 2018

algorithms. The experiments with each of the K-means, PSO, ABC, FFA, GOBL-GSO, FWA-ATFmethods are repeated for 51 times for each of the MR slices of both patients. The means and standard deviations of DB-Index values over 51 independent runs for each MR slice of the first and second patient are given in Table 1 and Table 2 respectively. From the results in Table, it is observed that the K-means performs better than SI based clustering methods for only two out of ten slices. Among the SI algorithms, ABC performs better clustering than others and K-means for four out of ten slices. From the results in Table 2, it is observed that the K-means method performs better than SI based clustering methods for only three out of ten slices. Among the SI algorithms, ABC performs better clustering than others. As ABC algorithm performs better than K-means and other SI algorithms for most of the MR slices of both patients, a pair wise comparative study of ABC with other algorithms has been carried out using Wilcoxon signed rank test[19]. The results of the aforementioned statistical test are given in Table 3 and 4 for thefirst and second patient respectively.

From the results in Table 3, it is observed that ABC algorithm performs better than K-means algorithm with a level of significance (α) =0.05. There are no statistically significant differences in the performances of ABC as well as other SI algorithms. From the results in Table 4, it is observed that ABC algorithm statistically performs better than PSO algorithm with significance level (α) =0.05. There are no statistically significant differences in the performances of ABC and other SI algorithms as well as K-means algorithm for the second patient. Computational complexities are given in Table 5 and Table 6 for the first and second patient respectively.PSO and FWA-ATF take lower computational time than other SI algorithms.

From the analysis of the above results, it is observed that all the SI based clustering algorithms perform better than K-means algorithm for the first patient. For the second patient, all the SI based clustering algorithms except PSO based clustering algorithm performs statistically equivalent to K-means algorithm. PSO and FWA-ATF take the lower computational time for segmentation of the image than other SI algorithms. But, PSO performs poor in segmentation of the images of the second patient. Therefore, FWA-ATF based clustering algorithm is more effective and efficient in the segmentation of soft tissue sarcomas.

IV. Conclusions

In this article, the Swarm Intelligence based clustering algorithms are applied in segmentation of soft tissue sarcomas in MR images. The SI-based clustering techniques statistically perform better than K-means algorithms in segmentation of MR images of the first patient dataset. There is no statistical difference in the performance of SI-based clustering techniques except PSO based clustering technique and K-means algorithm in segmentation of MR images of second patient dataset.

Though the most of the SI based clustering algorithms provide good performance in segmentation of soft tissue sarcomas, some other healthy tissues are also segmented as sarcomas because they have the similar intensity levels to sarcomas in MR images. The experimental results has shown that the SI-based clustering techniqueswould be useful for assisting radiologistsin the diagnosis of soft tissue sarcomas in MR images. In this work, we have used only gray levelfeature for MR image segmentation. Spatial or texture information can be used as features for better segmentation of soft tissue sarcomas in MR images.

Table 1: Mean And Standard Deviation Of Db-Index Over 51 Independent Runs For Each Image Of Patient-1.

MRI#	K-means		PSO		ABC		FFA		GOBL-GSO		FWA-ATF	
	Means	Std	Means	Std	Means	Std	Means	Std	Means	Std	Means	Std
1	0.1039	0.0002	0.1010	0.0028	0.0974	0.0006	0.0986	0.0017	0.0975	0.0008	0.0978	0.0002
2	0.1121	0.0088	0.1066	0.0020	0.1068	0.0004	0.1070	0.0010	0.1070	0.0006	0.1068	0.0003
3	0.1028	0.0011	0.1081	0.0037	0.1088	0.0003	0.1098	0.0016	0.1099	0.0012	0.1095	0.0006
4	0.1075	0.0016	0.1046	0.0054	0.1047	0.0016	0.1027	0.0039	0.1036	0.0025	0.1041	0.0022
5	0.1089	0.0066	0.1051	0.0011	0.1045	0.0004	0.1051	0.0006	0.1049	0.0004	0.1047	0.0002
6	0.1291	0.0115	0.1057	0.0027	0.1072	0.0006	0.1069	0.0013	0.1070	0.0011	0.1072	0.0002
7	0.1091	0.0020	0.1133	0.0030	0.1137	0.0001	0.1130	0.0018	0.1137	0.0006	0.1135	0.0006
8	0.1043	0.0005	0.0987	0.0046	0.0955	0.0003	0.0967	0.0021	0.0961	0.0008	0.0958	0.0002
9	0.1311	0.0011	0.1101	0.0044	0.1091	0.0026	0.1082	0.0015	0.1084	0.0018	0.1079	0.0018
10	0.1080	0.0062	0.1016	0.0014	0.1012	0.0004	0.1015	0.0009	0.1012	0.0004	0.1012	0.0004

Received: 20 September 2018/Accepted: 15 October 2018

Table 2: Mean and Standard Deviation of Db-Index Over 51 Independent Runs For Each Image of Patient-2

MRI#	K-means		PSO		ABC		FFA		GOBL-GSO		FWA-ATF	
	Means	**Std**	**Means**	**Std**	**Means**	**Std**	**Means**	**Std**	**Means**	**Std**	**Means**	**Std**
1	0.0986	0.0003	0.0901	0.0014	0.0886	0.0002	0.0892	0.0009	0.0890	0.0007	0.0888	0.0003
2	0.0819	0.0000	0.0902	0.0081	0.0812	0.0001	0.0834	0.0038	0.0810	0.0008	0.0812	0.0002
3	0.0872	0.0001	0.0940	0.0034	0.0879	0.0004	0.0912	0.0037	0.0889	0.0019	0.0890	0.0018
4	0.1004	0.0002	0.1133	0.0083	0.1059	0.0002	0.1027	0.0051	0.1031	0.0051	0.1021	0.0034
5	0.1079	0.0202	0.0939	0.0036	0.0898	0.0002	0.0923	0.0012	0.0909	0.0008	0.0910	0.0011
6	0.1156	0.0008	0.0947	0.0031	0.0932	0.0008	0.0944	0.0025	0.0938	0.0015	0.0942	0.0015
7	0.0930	0.0002	0.0913	0.0036	0.0873	0.0001	0.0885	0.0017	0.0875	0.0003	0.0873	0.0001
8	0.0839	0.0000	0.0945	0.0094	0.0844	0.0003	0.0876	0.0055	0.0851	0.0025	0.0866	0.0043
9	0.0931	0.0000	0.1005	0.0063	0.0928	0.0002	0.0950	0.0039	0.0934	0.0011	0.0935	0.0012
10	0.0990	0.0002	0.1012	0.0072	0.0939	0.0000	0.0949	0.0011	0.0944	0.0006	0.0940	0.0004

Table 3: Wilcox on Signed Ranks Test Statistics for the First Patient. R^+: Sum of Positive Ranks, r^-: Sum of Negative Ranks.

Sl. No.	Comparison	R^+	R^-	Z	p (2-tailed)
1	ABC *vs* K-means	47.00	8.00	-1.988	0.047
2	ABC *vs*PSO	34.50	20.50	-0.714	0.475
3	ABC *vs*FFA	31.50	23.50	-0.408	0.683
4	ABC *vs*GOBL-GSO	20.00	16.00	-0.281	0.779
5	ABC *vs*FWA-ATF	14.50	13.50	-0.085	0.933

Table 4: Wilcox on Signed Ranks Test Statistics for the Second Patient. R^+: Sum of Positive Ranks, r^-: Sum of Negative Ranks

Sl. No.	Comparison	R^+	R^-	Z	p (2-tailed)
1	ABC *vs* K-means	43.50	11.50	-1.632	0.1027
2	ABC *vs*PSO	55.00	0.00	-2.805	0.005
3	ABC *vs*FFA	46.50	8.50	-1.940	0.052
4	ABC *vs*GOBL-GSO	43.50	11.50	-1.633	0.1025
5	ABC *vs*FWA-ATF	28.00	8.00	-1.400	0.161

Table 5: Mean Computational Complexities (In Seconds) For the First Patient

Kmeans	PSO	ABC	FFA	GOBL-GSO	FWA-ATF
4.36	23.97	250.49	2524.07	386.37	23.72
4.16	23.73	236.57	2443.26	380.88	23.61
4.35	23.96	230.38	2448.21	380.80	24.39
4.27	23.94	230.27	2439.62	380.42	24.37
4.06	23.91	229.19	2430.82	379.97	23.63
4.03	23.65	239.53	2421.06	384.56	23.51
4.65	23.99	231.34	2435.17	391.52	24.02
4.37	23.96	229.49	2435.70	379.69	23.63
4.04	23.93	233.24	2448.97	381.60	23.70
4.02	23.85	233.28	2444.13	379.66	23.69

Received: 20 September 2018/Accepted: 15 October 2018

Table 6: Mean Computational Complexities (In Seconds) For The Second Patient

Kmeans	PSO	ABC	FFA	GOBL-GSO	FWA-ATF
3.84	23.76	235.42	1467.43	382.00	23.57
3.85	23.73	241.47	1452.66	382.60	23.42
3.87	23.77	239.30	1476.42	381.95	23.68
3.83	23.75	240.26	1464.41	385.84	23.85
3.83	23.75	238.06	1457.30	382.76	23.83
4.08	23.62	238.71	1445.53	381.66	23.70
3.90	23.74	237.48	1465.30	382.34	23.78
3.82	23.72	239.19	1448.40	383.68	23.87
3.87	23.78	239.07	2447.82	383.27	24.14
3.76	23.83	240.44	2441.73	384.57	24.15

References

[1] Demetri, G.D., Baker, L.H., Benjamin, R.S., Casper, E.S., Conrad, E.U., D'Amato, G.Z., DeLaney, T.F., Ettinger, D.S., Heck, R., Heslin, M.J. and Hutchinson, R.J. Soft tissue sarcoma: Clinical practice guidelines in oncology™. *JNCCN Journal of the National Comprehensive Cancer Network* **5** (4) (2007) 364-399.

[2] Clark, M.A., Fisher, C., Judson, I. and Thomas, J.M. Soft-tissue sarcomas in adults. *New England Journal of Medicine* **353** (7) (2005) 701-711.

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

[4] Farhidzadeh, H., Chaudhury, B., Zhou, M., Goldgof, D.B., Hall, L.O., Gatenby, R.A., Gillies, R.J. and Raghavan, M. Prediction of treatment outcome in soft tissue sarcoma based on radiologically defined habitats. In *Medical Imaging Computer-Aided Diagnosis*, 2015,1-5.

[5] Farhidzadeh, H., Zhou, M., Goldgof, D.B., Hall, L.O., Raghavan, M. and Gatenby, R.A. Prediction of treatment response and metastatic disease in soft tissue sarcoma. In *Medical Imaging 2014: Computer-Aided Diagnosis*, 2014, 1-6.

[6] Farhidzadeh, H., Goldgof, D.B., Hall, L.O., Gatenby, R.A., Gillies, R.J. and Raghavan, M. Texture feature analysis to predict metastatic and necrotic soft tissue sarcomas. *IEEE International Conference on Systems, Man, and Cybernetics*, 2015, 2798-2802

[7] Vallières, M., Freeman, C.R., Skamene, S.R. and El Naqa, I. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Physics in Medicine & Biology* **60** (14) (2015) 5471--5496

[8] Glass, J.O. and Reddick, W.E. Hybrid artificial neural network segmentation and classification of dynamic contrast-enhanced MR imaging (DEMRI) of osteosarcoma. *Magnetic resonance imaging* **16** (9, (1998) 1075-1083.

[9] Monsky, W.L., Jin, B., Molloy, C., Canter, R.J., Li, C.S., Lin, T.C., Borys, D., Mack, W., Kim, I., Buonocore, M.H. and Chaudhari, A.J. Semi-automated volumetric quantification of tumor necrosis in soft tissue sarcoma using contrast-enhanced MRI. *Anticancer research* **32** (11) (2012) 4951-4961.

[10] El-Dahshan, E.S.A., Mohsen, H.M., Revett, K. and Salem, A.B.M. Computer-aided diagnosis of human brain tumor through MRI: A survey and a new algorithm. *Expert systems with Applications* **41**(11) (2014) 5526-5545.

[11] Mitra, S. and Shankar, B.U. Medical image analysis for cancer management in natural computing framework. *Information Sciences*, 2015, 111-131.

[12] MacQueen, J. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* **1 (**14) (1967) 281-297.

[13] Omran, M., Engelbrecht, A.P. and Salman, A. Particle swarm optimization method for image clustering. *International Journal of Pattern Recognition and Artificial Intelligence* **19** (03) (2005) 297-321.

[14] Hancer, E., Ozturk, C. and Karaboga, D. Extraction of brain tumors from MRI images with artificial bee colony based segmentation methodology. *IEEE* In *International Conference on Electrical and Electronics Engineering (ELECO)*, 2013, 516-520.

[15] Manna, P. and Si, T. Brain mri segmentation for lesion detection using clustering with fire-fly algorithm. In *Artificial Intelligence and Evolutionary Computations in Engineering Systems*, 2016, 1347-1355.

Received: 20 September 2018/Accepted: 15 October 2018

[16] Si, T., De, A. and Bhattacharjee, A.K. MRI brain lesion segmentation using generalized opposition-based glowworm swarm optimization. *International Journal of Wavelets, Multiresolution and Information Processing* **14** (05) (2016) 1-29.
[17] Misra, P.R. and Si, T. mage segmentation using clustering with fireworks algorithm. *IEEE In International Conference on Intelligent Systems and Control (ISCO)*, 2017, 97-102.
[18] Davies, D.L. and Bouldin, D.W. A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence* **1** (2) (1979) 224-227.
[19] Derrac, J., García, S., Molina, D. and Herrera, F. A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. *Swarm and Evolutionary Computation* **1** (1) (2011) 3-18.
[20] Vallières, M., Freeman, C.R., Skamene, S.R. and El Naqa, I. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Physics in Medicine & Biology* **60** (14) (2015) 5471.
[21] Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M. and Tarbox, L. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *Journal of digital imaging* **26** (6) (2013) 1045-1057.

Received: 20 September 2018/Accepted: 15 October 2018