IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-6, Issue-7 E-ISSN: 2347-2693

MRI Image Analysis based on Reverse Classification Accuracy Method

Varsha E. Jaware[1*], Rajesh H. Kulkarni[2]

[1]Department of Computer Engineering, Rajshree Shahu School of Engineering and Research, JSPM NTC, Pune, INDIA
[2]Department of Computer Engineering, Rajshree Shahu School of Engineering and Research JSPM NTC, Pune, INDIA

Available online at: www.ijcseonline.org

Accepted: 21/Jul/2018, Published: 31/Jul/2018

Abstract — Magnetic Resonance Imaging (MRI) scan are the most widely used radiographic techniques in diagnosis, clinical studies and treatment planning. It is important to be able to detect when an automatic segmentation method fails to avoid inclusion of wrong measurements into subsequent analysis which could otherwise lead to incorrect conclusion. Sometimes due to absence of Ground Truth (manually labelled) images it is difficult to detect the failure of automatic segmentation methods. Before deployment, performance is quantised using different metrics for which the predicted segmentation is compared to a reference segmentation also known as Ground Truth (GT), which is manually obtained by an expert. In some exceptional cases it becomes difficult to know about its real performance after deployment when a reference is unavailable. To that end, this paper aims to develop an improved and advanced technique of Reverse Classification Accuracy (RCA) on new data which enables us to discriminate between the successful and failed cases. The ideal concept is that to rank the 'best' segmentation results in the database without knowing the manual label. Then 'match' the rank between the prediction and the truth image saved in database. Further, for correctly and accurately segmented and classified brain MRI images, diseases are being detected using Random forest algorithm and Deep Learning.

Keywords —Machine learning, Deep Learning, image Segmentation, MRI images, classification, performance evaluation

I. INTRODUCTION

Image Segmentation is an important factor for analysis which extracts clinical useful information from the segmented images. Image segmentation is the process of partitioning a digital image into multiple segments which is typically used to locate objects and boundaries (lines, curves, etc.) in images.

The field of medical imaging has undergone revolutionary advances over of the past 2 decades. New medical imaging technologies have provided physician's powerful, non-invasive techniques to probe the structure, operate and pathology of the physique. Segmentation among the particles of the varied components is extremely vital to medical call. Despite of the significant advancement in this field, medical image segmentation remains an unresolved problem. The level of subdivision in segmentation depends on the problem being solved. That is, segmentation should stop when the objects of interest in an application have been isolated.

With increasing use of Magnetic Resonance (MR) imaging scan for diagnosis, treatment planning and clinical studies, it has become almost compulsory to use computers to assist radiological experts in clinical diagnosis and treatment planning. Machine Learning classification has been the most important computational development in the last years to satisfy the primary need of clinicians for automatic early diagnosis. A two-class classification model for grouping and linear classifier to combine these features was proposed in [14]. To demonstrate the power of the classification model, a simple algorithm was used to randomly search for good segmentations. In cases of new data, it became very difficult to predict the real performance of a particular segmentation method when no reference image is available. RCA framework method was used to predict segmentation accuracy in such cases of absence of GT. Objective of this study is to find out how RCA model aims to answer the question on how to estimate the performance of models in cases where ground truth is not available [1].

Characterizing the performance of image segmentation approaches has been a persistent challenge. Performance analysis is plays a vital role since segmentation algorithms often have limited accuracy and precision. Interactive drawing of the desired segmentation by human raters has always been an acceptable approach, and yet suffers from intra-rater and inter-rater variability. Automated algorithms have been sought in order to remove the variability introduced by raters, but such algorithms must be assessed to check their endurance whether they are suitable for the task or not. The performance of raters (human or algorithmic) generating segmentations of medical images has been difficult to quantify because of the difficulty of obtaining or estimating a known true segmentation for clinical data. An example of segmented image is shown in Fig.1

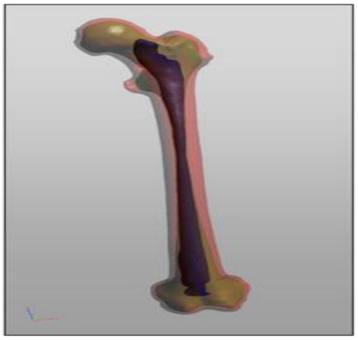

Fig.1 A Segmented image of Femur Bone

The trend towards large-scale studies including population imaging poses a new challenge in terms of quality control. This is a particular issue when automatic processing tool such as Image segmentation methods are employed to derive quantitative measures or bio markers for further analyses. However it is important to detect when an automatic method fails to avoid inclusion of wrong results which further lead to incorrect conclusions. In order to overcome this challenge, reverse classification accuracy approach is used for predicting segmentation quality. Traditionally segmentation performance was evaluated by selection of a random data set and obtaining a manual expert segmentation for it which was then compared to the automatic one. In order to solve segmentation problems, different methods were proposed based on statistical models [2], multi-atlas label propagation [3] and supervised classification [4].

On deploying segmentation method there is hardly any idea to predict its real performance. Assessing the performance using traditional evaluation measures is a challenging problem when no GT is available for comparison. Perhaps, it is difficult to assess the accuracy level on clinical data [15] and thus find out the failure of automatic segmentation method. Particularly when segmentation is an intermediate step in large scale automated analysis where no visual quality control is present. This plays a vital role in large scale studies like the UK Biobank Imaging Study [16] where automated methods are applied on thousands of images, and segmentation is used for further statistical population analysis.

In clinical routine the interest is in per case performance and wants to evaluate when the automated method fails. The problem faced is that the performance of a method is substantially different on clinical data and is generally lower than what is found through cross validation on annotated data carried out previously for many reasons. Firstly incremental method development for training the annotated database is normally used for model selection and fine tuning of hyper parameters. Secondly the clinical data is different due to the varying imaging protocols or artefacts caused by the pathology. This paper presents an approach where a solution to a framework for predicting the real performance of deployed segmentation method on a per case basis of a new image in the absence of GT.

Motivation to the above observations, we in this paper improvised RCA model for reverse testing and extending further to detect diseases in correctly classified brain images.

This paper mainly contributes the following taks : *

* We classify the segmented image in a unique reverse way by using Deep Learning technique, when no ground truth is available.

* We will detect the quality and performance of segmentation method by using ASD and HD distance evaluation metrics and also improve precision along with traditionally used Dice Similarity coefficient (DSC) an overlap based measure.

* We propose to apply Deep Learning to RCA classifier and obtain good segmentation quality results

* Further extending to correctly classified brain images we will detect whether there is a possibility to develop diseases for that particular patient

II. RELATED WORK

In many domains from graphics, remote sensing to marketing Strategies, it has become difficult to retrieve an objective performance evaluation without GT. The lack-of-label problem has been tackled by exploiting transfer learning in [5] using a reverse validation procedure when the number of labeled data is limited in the general machine learning domain. The basic idea of reverse validation [5] is based on reverse testing [6] new proposed classifier is initially trained on prediction of the test data and later evaluated on Training data. This thought of reverse testing is closely related to the approach of RCA in [1].

The segmentation performance is evaluated by several tasks like separating the perceptual salient structures [7], automatically generating semantic GT [8], [9] or by observing at contextual properties [10].A method for analyzing objective metrics like precision and recall with no GT is proposed but it failed with data sets that had partial GT as probabilistic model approach is used.

System proposed by [11] used region-correlation matrix to quantify the performance of various segmentation algorithms. The drawbacks found with this system were that there was no independence for evaluation of the segmentation performance considering a particular method for an image.

Meta-evaluation framework, which is a standalone method where image features are used to provide ranking of different methods [17] used in a machine learning setting that provides a ranking of different methods. Standalone methods have the advantage that they do not require a manually segmented reference image for comparison and can therefore be used for real-time evaluation. But it faced a problem with the estimation of segmentation performance of an individual image. An effort to reckon objective metrics like precision and recall with missing GT is proposed by [18]. The drawback with it was that it could be used for data sets with partial GT since under the same assumptions it applies a probabilistic model.

Unsupervised methods in [19], [20] aim to obtain the segmentation accuracy directly from the images and label maps like geometrical and information-theoretic features. It enables the quantification of the quality of an image segmentation result. These evaluation criteria reckon some statistics for each region or class in a segmentation result. Such an evaluation criterion can be useful for different applications like the comparison of segmentation results, the automatic choice of the best fitted parameters of a segmentation method for a given image, or the definition of new segmentation methods by optimization. But the application in medical setting is unclear as unsupervised methods are applied to scenarios where the main motto of segmentation is to yield visually consistent results that are meaningful to a human observer.

Among all ensembles approaches Random Forest (RF) [11] produced the best accuracies in many scientific fields. Random forests grow many classification trees. To classify a new object from an input vector, put the input vector down each of the trees in the forest. Each tree gives a classification and the tree "votes "for that class. The forest chooses the classification having the most votes (over all the trees in the forest).

III. EXISTING SYSTEM

A system framework aimed to asses the performance of the segmented images in a reverse manner has helped to solve cases where there is no reference available to comapare with. RCA is applied to estimate the segmentation quality by using one of the specific segmentation methods like Multi Atlas or Random Forest.RCA model was applied with Single Atlas, Constrained CNN and Atlas Forest. But among that algorithm Constrained CNN failed to give a high accuracy Single Atlas and Atlas Forest. This model worked well on large body organs like brain, stomach, liver, pelvis and but estimated less accuracy for smaller organs like spleen, adrenal glands and clavicles. There was a need to improvise the method not only for the purpose of predicting accuracy but also for further detection of diseases. System has two phases, mainly segmentation phase and then comparing it with the reference database for predicting the accuracy in a reverse manner. The supposition was that the RCA classifier which was trained using estimated segmentation would pretend to be pseudo GT, worked well only if the quality of segmentation was high for a newly obtained image. Similarly, if the quality of segmentation revealed, was poor then it would poorly perform on the reference images. The accuracy was quantified using the traditional evaluation metric, Dice Similarity coefficient. Previously machine learning data relied on single classifier algorithms like Support vector Machine (SVM), Naive Bayes, or Linear Discriminant Analyses. But in the last few years, ensembles algorithms [11] have resulted in exploring another reliable alternative option to single classifiers which gave better performance than the latter one.

IV. PROBLEM STATEMENT

Predicting the performance of a segmentation method on new data, estimation of the quality of segmented image and detection of the failure cases when there is no Ground Truth available to compare with is a challenging problem in medical image analysis. In addition to it detection of disease is done for better and improved analysis. Therefore, a new way for classification in a reverse way is introduced in the RCA Model.

V. PROPOSED SYSTEM

In the context of such large scale analysis improvised automatic quality control is a necessity in RCA. In this paper we propose an improvised RCA model which predicts more high accuracy than before, by the use of Random Forest classification algorithm and further detect AD in correctly classified images of the particular patient.

A. System Overview

Considering the scenario that in cases of absence of GT images, it becomes difficult to analyze the segmentation performance. It is also necessary to check when the automatic segmentation method fails, thus its detection is also necessary. Along with it on obtaining correctly classified brain images , further detection of AD, if present is found out. Fig.2 shows the detailed architecture and flow of the process.

B. Mathematical model

Equation of a proxy measure for predicting the segmentation accuracy is as follows:

Where is any evaluation metric, such as DSC,ASD or HD

*SI denotes the predicted segmentation that here acts as pseudo GT

*I be an image that has been segmented by any segmentation method

*Segmentation Function FI, SI (J) = SJ

*Image J which produces a segmentation SJ

$$\rho.\ (SI) = \max_{1 \le k \le m} \rho(FI, SI\ (Jk),\ S\mathrm{GT}\ Jk)$$

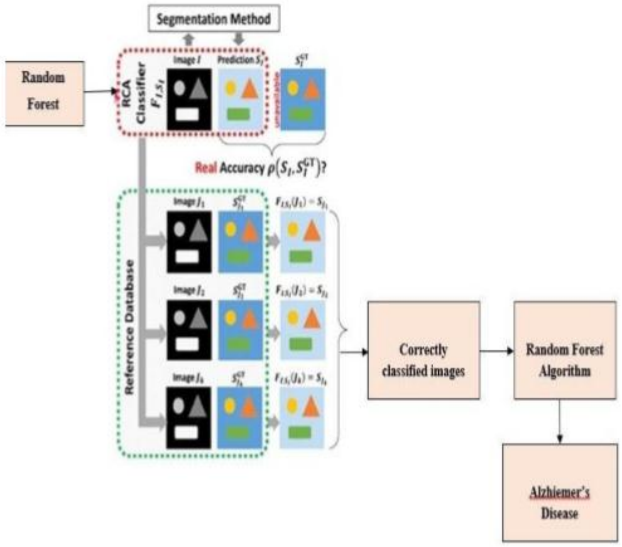

Fig 1 : Proposed System Architecture

C. Model Description

Segmentation phase: In this step, segmentation of MRI or CT scan images of organs like brain, liver, lung, etc. is done using an improvised algorithm named as Random Forest. Images are obtained from online or real time datasets.

RCA Model: The input received from segmentation phase is then applied to RCA model for classification and predicting its performance by comparing it with the reference database.

Reference Database: For storing images we have two partitions of database named as Training and Testing database. Absence of GT is only for Testing database (which we wish to predict using the RCA). To get the segmentation on testing database, we used reference (Training) database to train a classifier. This training database has GT, which we used to train and segment testing database - and also to predict the segmentation accuracy using improved RCA model. We have divided database in two portions by following 70/30 rule i.e. 70% for training and 30% for testing.

Disease Detection phase: Further, after correctly classifying brain images from the RCA model, the presence of disease is being detected.

D. Software Requirement Specification

Software Requirements:

* Operating System: Windows 7 or 10 * Programming Language: Python

* Datasets: MRI (Brain, Liver)

Hardware Requirements:

* Processor: Intel(R) Core(TM) 2 Duo CPU 1.1 GHz

* RAM: 2 GB

E. Steps to be Carried

Step 1: Initially segment the MRI scan images for which GT is absent.

Step 2: The segmented image obtained from previous step is then applied to RCA Model.

Step 3: RCA Model then Classifies those segmented images and then predicts its accuracy.

Step 4: Further correctly classified Brain images are used for disease detection in them.

Step 5: Using Random Forest algorithm the presence of disease is detected.

F. Input and Output :

Input:

MRI images

Output:

Classified Brain MRI images with disease detection.

G. Algorithm

Random Forest method is used for Segmentation Phase. RCA model includes CNN methods for classification

a) Random Forest Algorithm:

[A] Random Forest creation pseudocode:

1. Randomly select K features from total m features where k << m.
2. Among the K features, calculate the node d using the best split point.
3. Split the node into daughter nodes using the best split.
4. Repeat the a to c steps until l number of nodes has been reached
5. Build forest by repeating steps a to d for n number times to create n number of trees.

[B] Random forest prediction pseudocode:

1. Takes the test features and use the rules of each randomly created decision tree to predict the outcome and stores the predicted outcome (target) .
2. Calculate the votes for each predicted target.
3. Consider the high voted predicted target as the final prediction from the random forest algorithm.

Each tree is grown as follows:

1) If the number of cases in the training set is N, sample N cases at random but with replacement, from the original data. This sample will be the training set for growing the tree.

2) If there are M input variables, a number m is specified such that at each node, m variables are selected at random out of the M andthe best split on these m is used to split the node. The value of m is held constant during the forest growing.

3) Each tree is grown to the largest extent possible. There is no pruning.

b) Deep Learning:

CNNs are also used as one of the RCA classifiers. A Deep Medic is utilized for a 3D CNN architecture for automatic segmentation [21]. The architecture is computationally efficient as it can handle large image context by using a dual pathway for multi-scale processing. CNNs are able to learn highly complex and discriminative data associations between input data and target output. The architecture of the network is defined by the number of layers and the number of activation functions in each layer. In CNNs, each activation function corresponds to a learned convolutional filter, and each filter produces a feature map (FM) by convolving the outputs of the previous layer. Through the sequential application of many convolutions, highly complex features are learned that are then used to produce voxel-wise predictions at the final, fully-connected layer.

CNNs are a type of deep learning approach which normally requires large amounts of training data in order to perform well due to the thousands (or millions) of parameters corresponding to the weights of the filters. To be able to act as a RCA classifier that is trained on a single image, a specialized architecture is required. The number of FMs is reduced in each layer by one third compared to the default setting of Deep Medic. The feature maps in the last fully connected layers from 150 to 45 are also cut. The network preserves its capability to see large image context as the size of the receptive field remains unchanged by reducing the feature maps without changing the architecture in terms of number of layers. With less number of filters, the number of parameters is substantially decreased, that leads to faster computations, but more importantly, reduces overfitting when trained on a single image. The original input image is divided into 3D patches that are then sampled during training using backpropagation and batch normalization where the training is performed in a patch-wise manner.

H. Performance Metrics for Prediction Accuracy:

The Dices similarity coefficient is the most widely used measure for evaluating segmentation performance , and the main results in this study focused on how well the DSC, predicted good results using RCA framework. In order to quantify prediction accuracy, three different measures, namely the correlation between predicted and real DSC, the mean absolute error (MAE), and classification accuracy were incorporated in this model. Arguably, the most important measure for direct evaluation of how well RCA works is the MAE directly reveals how close the predicted DSC is to the real one. Correlation is of great importance, as it conveyed a relation between predicted and real scores.

E. Segmentation Failure Detection In clinical measures it is of great importance to be able to detect when an automated method fails. Low real DSC scores if obtained are correctly predicted and failed segmentations are identified from the score. It is important to exclude failed segmentations from the subsequent analysis as it leads to wrong results or conclusions.

I. *Summary:*

An appealing property of the proposed framework is that unlike the supervised methods no training data is required that captures examples of good and bad segmentations. Instead, RCA simply relies on the availability of a reference database with available GT segmentations. The drawback, however, is that assumption of a linear relationship between predicted and real scores which should be close to an identity mapping, something only found in the case of using Single-Atlas label propagation. In the case of off-diagonal correlation, as for example found for Atlas Forests, an extension to RCA could be considered where the predictions are calibrated. This, however, requires training data from which a regression function could be learned.

VI. PEFORMANCE AND RESULT

The final results are represented in gist in Tab.1, where quantitative analysis of predicted accuracy is shown. We have used Single Atlas forest for Segmentation purpose and Random Forest as an algorithm for RCA analysis. The correlation achieved between the DSC's overlap based score of predicted and real values is good. We improvise the Distance based metric in this system as it resulted in low correlation, high value Mean Absolute Error (MAE) and low accuracy rate of classification in the previous system. Distance based errors are unbounded hence a threshold value is being set for HD and ASD which lays the boundary reference values for categorizing error rate. For HD the boundary reference values are [0,10]for good segmentation quality, [10,60] for medium segmentation quality, [60,150] for bad segmentation quality. ASD's reference boundary range is as follows [0,2], [2,5], [5,10] for good, medium and bad quality segmentation .For this experiment Single Atlas Forest or Atlas forest are used for segmentation and Random Forests as RCA classifier. The results are encapsulated in Tab I using Random Forest (proposed) method.

Table 9.1: Predicting Different Segmentation Metrics for Random Forest(proposed)

Metric	Correlation(%)	MAE(%)	Accuracy(%)
DSC	90	10	80
HD	25	35	40
ASD	25	3	40

The results are summarized in Tab II using Atlas Forest (existing).

Table 9.2: Predicting Different Segmentation Metrics for Atlas Forest (existing)

Metric	Correlation(%)	MAE(%)	Accuracy(%)
DSC	88	12	78
HD	18	41	38
ASD	22	4	34

VII. CONCLUSION

The optimal solution of the RCA analysis reveals the possibility of precisely estimating the segmentation quality of medical images on the lack of Ground Truth and further detects whether there is a presence of any disease or not in the correctly classified image. Using new RCA model, the prediction accuracy is improvised that proves to be a boon for integrating it into an automatic processing pipelines when used in clinical routine study. It is of vital importance in case of large scale analysis to study the detection of failed segmentation images as there is no feasibility in employing manually assisted quality control along with visual inspection. Further detection of disease in such cases is also done that proves to be a time saving analysis. In future the proposed system may try to find out liver and other organs disorders after correctly classifying liver images from RCA model.

ACKNOWLEDGMENT

I express true sense of gratitude, sincere and heartfelt thanks to my guide Prof. R H. Kulkarni, HOD of Computer department for his invaluable co-operation and guidance that he gave me throughout my research, for inspiring and providing me all the lab facilities, which made this research, very convenient and easy.

REFERENCES

[1] Vanya V. Valindria, Ioannis Lavdas, Wenjia Bai, Konstantinos Kamnitsas, Eric O. Aboagye,Andrea G. Rockall, Daniel Rueckert, and Ben Glocker,*"Reverse Classificaton Accuracy: Predicting Segmentation Performance in the absence of Ground Truth,"* IEEE Transactions on Medical Imaging, Vol. 36, No. 8, 2017.

[2] T. Heimann and H.-P. Meinzer, *"Statistical shape models for 3D medical image segmentation: A review"* Med. Image Anal., vol. 13, no. 4, pp. 543-563, 2009.

[3] J. E. Iglesias and M. R. Sabuncu, *"Multi-atlas segmentation of biomedical images: A survey"* Med. Image Anal., vol. 24, no. 1, pp. 205-219,2015.

[4] E. Geremia et al., *"Classification forests for semantic segmentation of brain lesions in multi-channel MRI"* in Decision Forests for Computer Vision and Medical Image Analysis. London, U.K.: Springer, pp. 245260, 2013.

[5] E. Zhong, W. Fan, Q. Yang, O. Verscheure, and J. Ren, *"Cross validation framework to choose amongst models and datasets for transfer learning"* in Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases, pp. 547-562, 2010.

[6] W. Fan and I. Davidson, *"Reverse testing: An efficient framework to select amongst classifiers under sample selection bias,"* in Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, pp. 147-156, 2006.

[7] F. Ge, S. Wang, and T. Liu, "New benchmark for image segmentation evaluation," J. Electron. Image. vol. 16, no. 3, pp. 033011, 2007.

[8] L. Goldmann et al., *"Towards fully automatic image segmentation evaluation,"* in Proc. Int. Conf. Adv. Concepts Intell. Vis. Syst., pp. 566-577, 2008.

[9] H. Li, J. Cai, T. N. A. Nguyen, and J. Zheng, *"A benchmark for semantic image segmentation"*,in Proc. IEEE Int. Conf. Multimedia Expo, pp. 16,2013.

[10] K. Sikka and T. M. Deserno*"Comparison of algorithms for ultrasound image segmentation without ground truth*", Proc. SPIE, vol. 7627, pp. 76271C176271C9, 2010.

[11] Leo Breimen, *"Random Forest"*, 2001.

[12] K.Yogeswara Rao, M.James Stephen, D.Siva Phanindra,*"Classification Based Image Segmentation Approach",* IJCST, Vol 3, Issue 1, 2012.

[13] Xiaofeng Ren and Jitendra Malik,*"Learning A Classification Model for Segmentation".*

[14] Dr Stefan Teipel, Michel J Grothe, , Henryk Barthel,*"Multimodal imaging in Alzheimer's disease: validity and usefulness for early detection ",*Lanclet Neurology,2015.

[15] B. van Ginnekan, T. Heimann, M.Styner,*"3D Segmentation in the clinic: A Grand Challenge "*, in proc. MICCAI Workshop 3D Segmentation Clinic, pp. 7-15, 2007.

[16] C. Sudlow et al.,*"UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age ",* PLoS Med., vol. 12, no. 3, p.e1001779, 2015.

[17] H. Zhang, S. Cholleti, S.A. Goldman, J.E.Fritts,*"Meta-evaluation of Image Segmentation using Machine Learning ",*in Proc. IEEEE Comput. Soc. Conf. Compute. Vis. Pattern REcognit. (CVPR). vol. 1, pp. 1138-1145,2006.

[18] B.Lamiroy and T. Sun,*"Computing Precision and Recall with Missing or Uncertain Ground Truth",* in Graphics Recognition. New Trends and Challenges. Berlin, Germany: Springer- Verlag, pp. 149-162,2013

[19] S. Chabrier, B. Emile, C. Rosenberger, H. Laurent,*"Unsupervised Performance Evaluation of Image Segmentataion"*, EURASIP J. Appl.Signal Process., vol. 2006, pp. 217-229, 2006.

[20] H. Zhang, J. E. Fritts, and S. A. Goldman, *"Image Segmentation Evaluation: A survey of Unsupervised Methods",*Computer Vision and Image Understanding , vol. 110, no. 2, pp. 260-280, 2008.

[21] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, B. Glocker, *"Efficient Multi-scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation,"* Medical Image Analysis, 2016.

[22] W. Bai, W. Shi, D. P. ORegan, T. Tong, H. Wang, S. Jamil-Copley, N. S. Peters, and D. Rueckert,*"A Probabilistic Patch-Based Label Fusion Model for Multi-Atlas Segmentation with Registration Refinement: Application to Cardiac MR Images",* IEEE Transactions on Medical Imaging, vol. 32,no. 7, pp. 1302-1315, 2013

Authors Profile

Varsha E. Jaware, *has* received the ***Bachelor*** of Engineering Degree in Computer Science & Engineering from, GSMCOE, Pune University, India in year 2013. She is currently pursuing Master of Engineering from Rajshree Shahu School of Engineeringand Research, JSPM Pune,India.

Rajesh H. Kulkarni is currently working as HOD in the Department of Computer Engineering, Rajshree Shahu School of Engineering and Research, JSPM NTC, Pune, India. His research area includes Software Engineering.

