International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

# Image Segmentation Using Machine Learning for Multimodal Medical Images

[1]Dr. Kajal Chheda, [2]Akhilendra Pratap Singh, [3]Raman Verma, [4]Sunil M. P., [5]Shweta Madaan

**Submitted**: 08/01/2024 **Revised**: 14/02/2024 **Accepted**: 22/02/2024

**Abstract:** A basic assignment in medical imaging, the segmentation of images is essential for many therapeutic applications. Because of the inherent difficulty and variability in various types of imaging, separating multimodal medical pictures presents major hurdles. The research suggests a novel image segmentation method called bird swarm optimized random forest (BSORF) for multimodal medical concepts. The BSO algorithm is used in the proposed approach to enhance the feature selection procedure and make it possible to identify the best particular characteristics in multimodal clinical images. The RF method, renowned for its efficiency when processing complex data and categorization assignments, is then used with these chosen characteristics as input. Numerous tests were run on multimodal medical data to assess our strategy's performance. The results show that the suggested method outperforms current approaches regarding accurate segmentation.

***Keywords:*** *Image segmentation, multimodal medical images, therapeutic applications, birds swarm optimized random forest (BSORF)*

## 1. Introduction

The term "multimodal medical images" refers to various imaging techniques to record and examine different human body parts for diagnosis, therapy, and research. These methods integrate many imaging modalities, including ultrasound, positron emission tomography (PET), computed tomography (CT), magnetic resonance imaging (MRI), and others, to provide a more thorough and in-depth understanding of the internal structures and functions. [1]. Multimodal medical images offer enormous potential for individualized treatment planning in this era of customized healthcare. Medical professionals can improve treatment plans and provide individualized care by fusing imaging data with additional clinical data such as genomes and patient history. Incorporating image-guided therapies is also made easier by multimodal imaging, allowing precise targeting of sick tissues and reducing collateral harm [2].

A cutting-edge method for partitioning or classifying distinct regions or structures within medical images obtained through several imaging modalities is called image segmentation utilizing machine learning for multimodal medical images. Healthcare workers can now extract valuable information from large and varied imaging datasets, vital for medical diagnosis, treatment planning, and research [3]. Image segmentation is only one of the many computer vision tasks that machine learning algorithms, especially deep learning (DL) models, have proven remarkably successful at. Large datasets of medical pictures tagged by humans with the necessary regions or features of interest can be used to train these models. Based on the trends and attributes it extracts from the initial data, the algorithm for machine learning is trained to recognize and divide these regions into several categories [4]. Machine learning-based image segmentation has many uses in multimodal medical pictures. It can help accurately delineate organs, tumors, blood vessels, and other anatomical characteristics, aiding radiation therapy, treatment planning, and surgical guidance. Additionally, it can improve clinical decision-making by tracking treatment outcomes, assessing illness progression, and measuring disease severity. Additionally, the technique enables large-scale evaluation of multimodal datasets, opening up new opportunities for medical research [5].

The study aims to present a new technique for multimodal medical picture segmentation dubbed bird swarm optimized random forest (BSORF). The researchers' use of the BSO algorithm improves the feature selection process, which enables the recognition of the most useful features in multimodal clinical images. Using the chosen features as input, they then employ the random forest (RF) approach, renowned for its effectiveness in handling complex data and categorization jobs. The researchers ran several tests on various multimodal medical data to gauge how well their suggested strategy worked.

[1]*Department of ISME, ATLAS SkillTech University, Mumbai, Maharashtra, India, Email Id- kajal.chheda@atlasuniversity.edu.in, Orcid Id- 0000-0002-7803-3510*
[2]*Maharishi University of Information Technology, Lucknow, India -226036, Email Id- akhilendrasingh.muit@gmail.com, Orcid Id- 0000-0002-2780-1073*
[3]*Chitkara University, Rajpura, Punjab, India raman.verma.orp@chitkara.edu.in, https://orcid.org/0009-0006-5137-2260*
[4]*JAIN (Deemed-to-be University), Karnataka, India, Email Id- mp.sunil@jainuniversity.ac.in, Orcid Id- 0000-0002-7737-4145*
[5]*Vivekananda Global University, Jaipur shweta.madaan@vgu.ac.in, Orcid Id- 0009-0009-0396-0254*

The order in which the remaining information is delivered is shown below. Section 2 covers the literature review, while Section 3 discusses methodologies. Section 4 examines the findings, and Section 5 presents the conclusions.

## 2. Related Works

The article [6] gives young researchers a sane perspective on creating CNN models combined with medical imagery to detect diseases early. Different image processing methods that improve the presentation of medical images have been covered in the article. High-quality medical photos can help the DL model perform more accurately. Study [7] Convolution neural networks have recently made strides toward producing more accurate prediction outcomes for issues with segmenting medical images. They extract knowledge from highly skilled medical image segmentation networks to train another lightweight network efficiently. They illustrate it requires reasonably high operational speed and minimal storage utilization, and a lightweight network distillation by our method provides a non-negligible value.

In the research [8] to learn enough from the limited input, the proposed CNL-UNet features a pre-trained encoder that enhances transfer learning algorithms. As a result, it has a lightweight design and is less likely to overfit. The outcomes show that their suggested architecture outperforms most of the current networks. The aim of the study [9] was to highlight the cutting-edge approaches to medical image fusion, with a particular emphasis on the use of wavelet transforms for the fusion of medical images, along with the help of principal component analysis (PCA) and independent component analysis (ICA) approaches for data dimension reduction and demising. The article has also covered the use of numerous medical modalities to create an optimum method for medical picture fusion.

An approach for the registration and merging of multimodal medical images was presented in the study by [10]. The proposed method uses Principal Component Averaging (PCAv) and Discrete Wavelet Transform (DWT) for picture fusion and Multi-resolution Rigid Registration (MRR) for multimodal image registration. The suggested methodology offers improved image quality, more accurate results, and useful data for diagnosis. To further boost the efficiency of the image fusion process, a novel deep belief network-based architecture is provided in the investigation [11]. The machine learning model for picture fusion is built using a deep belief network model. Numerous tests show that the suggested strategy performs better than the state-of-the-art imagines fusion methods.

The study's [12] Authors proposed that HyperDenseNet, a 3-D fully convolutional neural network, to apply the idea of dense connectivity to multimodal segmentation problems: HyperDenseNet, a CNN with applications for segmenting brain tissue in multimodal MRI. The study revealed how HyperDenseNet might be used to solve difficult multimodal volumetric image segmentation difficulties.The goal of study [13] is to refer to various MRI images obtained under different conditions as multiple modalities. Generative adversarial networks (GANs), a DL model, are the foundation of the suggested approach. Experimental findings show that the recommended strategy outperforms the most recent cutting-edge techniques.

In the research [14], to improve the prediction of the diagnostic details, multimodal medical image fusion attempts to combine multisensory medical data into a single image. The article uses a convolution neural network (CNN) as the suggested method. The recommended method improves other cutting-edge techniques regarding outcomes, fusion quality, and computational performance. The study [15] gives a thorough topical assessment of DL-based medical Image segmentation. The study focuses on supervised and weakly supervised learning approaches. Deep neural network-based, fully automated medical image segmentation has been demonstrated to be extremely valuable.

## 3. Proposed Method

### 3.1 Dataset

Frontal Chest X-ray (CXR) pictures were created using an openly accessible dataset. MRI scans come from the steady-state Sunny Brook dataset. Forty-five patients are given free precision (SSFP) sequences. All patients got multiple-slice short axis (SAX) images; however, some also had two-chamber and four-chamber views. We only used the slices of the CT images that contained the heart during the chest CT scans we performed on 20 patients. The echo ultrasound data comprises four chamber B-mode views and Doppler from a partner hospital.

### 3.2 Preprocessing - wiener filter

The wiener filter is a widely used image processing technique for eliminating extraneous information from the Multimodal Medical image. Our approach converted the input image into a two-dimensional representation with pixels having values between 0 and 255. The Multimodal Medical image was further altered using the same method to execute matrix operations. Since the Wiener filter uses a linear estimation of the original image, it removed additive noise and lowered mean square error. Equation 1 and Equation 2 thereby calculated the mean and variance of each pixel.

$$\mu_{a,b} = \frac{1}{AB} \sum_{x,y \in N} P(x,y) \qquad (1)$$

$$\sigma^2_{a,b} = \frac{1}{AB} \sum_{x,y \in N} P(x,y)^2 - \mu^2 \qquad (2)$$

Here, $P$ is an input image, $x$ and $y$ are filtering pixels, N=X=3 are surrounding pixels, and $x$ and $y$ are placed in an $N \times X$ block.

$$P_{(a,b)} = \mu + \frac{\sigma^2 - n^2}{\sigma^2} ( P(x-y) - \mu ) \quad (3)$$

In Equation 3, where $n^2$ is the noise variance, changes according to the Wiener filter are made.

### 3.3 Bird swarm optimized random forest (BSORF)

BSO-RF is the integration of bird swarm optimization an random forest for multimodal medical image segmentation.

#### 3.3.1 Bird swarm optimization

BSA is a powerful optimization technique with attributes including an easy process, good expansibility, and more.

Let N simulated birds may fly and search for food. Suppose $Z_x^h ( x \in [1,2, \ldots .. N])$. Describe where the x th bird at h is located. These are some examples of the birds' behaviors:

$$Z_{x,y}^{h+1} = Z_{x,y}^{h} + (P_{x,y} - Z_{x,y}^{h}) * D * rand(0,1) + (H_{x,y} - Z_{x,y}^{h}) * S * rand(0,1) \quad (1)$$

The following are examples of vigilance behavior:

$$Z_{x,y}^{h+1} = Z_{x,y}^{h} + B\,(mean_y - Z_{x,y}^{h}) * rand(0,1) + B_2\,(P_{x,y} - Z_{x,y}^{h}) * rand(-1,1) \quad (2)$$

Where B1 and B2 have the following mathematical definitions:

$$B_1 = b_1 * exp\left[ - \frac{p_{Fit_x}}{sum\ Fit + \in} * N \right]$$

$$B_2 = b_2 * exp\ \left[\left[ \frac{p_{Fit_x} - p_{Fit_k}}{|p_{Fit_k} - p_{Fit_x}| + \in} \right] * \frac{N * p_{Fit_k}}{sum\ Fit + \in} \right]$$

The following sentence describes flight behavior:

$$Z_{x,y}^{h+1} = Z_{x,y}^{h} + randn(0,1) * Z_{x,y}^{h} \quad (3)$$

$$Z_{x,y}^{h+1} = Z_{x,y}^{h} + \left( Z_{k,y}^{h} - Z_{x,y}^{h} \right) * FL * randn(0,1) \quad (4)$$

Where FL is in [0, 2].

#### 3.3.2 Random forest

Segmenting chest multimodal medical images is critical for computer-aided evaluation and therapy planning of chest -related disorders. Using the random forest algorithm is one way to carry out this segmentation.A supervised machine learning approach called random forest makes predictions by combining the outputs of various decision trees. The algorithm can be developed with labeled images where every pixel is identified as either a member of the area surrounding the chest s or not in segmenting chest multimodal images. A training dataset is first created, which includes ground truth segmentation masks for chest multimodal images. The image are usually preprocessed to highlight the chest architecture and eliminate noise or artifacts. The qualities of various regions can be represented by features that can be retrieved from the images, such as intensity values, texture, and form descriptors.

The random forest algorithm is trained on this dataset, and each decision tree is built by randomly choosing a portion of the training data and features. The link between the extracted features and the matching class labels is taught to the trees through training. The algorithm divides the data iteratively depending on various feature thresholds during training to develop a set of classification rules. The random forest model can segment new chest multimodal medical images after training. In the same way as the training photos, the image is initially preprocessed. Then, using the learned rules, the trained random forest is used to categorize each pixel in the image. This process is often carried out pixel-by-pixel to make a prediction, looking at each pixel's immediate surroundings.

Following segmentation, post-processing procedures can be used to clean up the data and eliminate any potential artifacts or misclassifications. These steps may involve morphological operations such as erosion and dilation to adjust the boundaries and ensure precise and seamless chest segmentation. The powerful machine learning method, random forest, can be applied to segment multimodal medical images. The system learns to categorize pixels in multimodal medical images by training on a labeled dataset. This method can help doctors better understand chest -related disorders and plan their diagnosis and treatment.

### 3.5 Classification using Support Vector Machine

We used SVM to categorize the Multimodal Medical Image as normal or abnormal. SVM is a methodical technique for two classroom issues. Considering how well the SVM classifier performs recognizing patterns and image processing applications, it is used in many study fields. SVM is most commonly applied to issues involving a limited training dataset and a large feature space. SVM requires two stages: training and testing, just like neural networks. The SVM's algorithm for learning can be trained by feeding its features as input. During exercise, the SVM determines the proper margin between the two classes. Based on the class they belong to, parts are given names[19].

The aggregate amount of neurons chosen for every problem and the existence of regional minimum constitute merely two of the numerous drawbacks of ANN. SVM does not occupy any local minima, and it solves the neuron selection issue by introducing the hyperplane concept. Our SVM uses an RBF kernel to translate the input information into space with greater dimensions. The pattern vector lies most near the decision boundary and is utilized for the hyperplane linear classification algorithm in this altered location.

Let's say that the $D$-dimensional inputs ax *(x=1,…D)* are either from Class I or II and the corresponding labels are bx=1 for classes one and two.

$$C(Y) = M^h X + A \qquad (1)$$

A - is a scalar, and M - is an m-dimensional vector. The dividing hyperplane meets this need.

$$Z_x\,(M^h X_x + \mathrm{b}) \geq 1 \quad for\ x = 1, \ldots .D \qquad (2)$$

The margin measures the separation between the practice data closest to the hyperplane and its edge, *C(Y) = 0*. The ideal separating hyperplane is the *C(Y) = 0* with the largest margin.

## 4. Result

This section evaluates the efficacy of the recommended and existing approaches. The variables are Dice Similarity Coefficient (DSC), precision, accuracy, and Jaccard Index (JI). CMSA [16] and MSMCNN [17] are the current processes.

The segmentation findings' overall correctness is gauged by accuracy. The ratio of true negatives and true positives to all the pixels or voxels in the image is used to compute it. However, if the classes are unbalanced, accuracy could be deceptive. The resulting accuracy is shown in Figure 1. Comparatively, it demonstrates that the proposed technique's value (proposed method 97%) is higher than the existing methods (CMSA 85% and MSMCNN 92%).

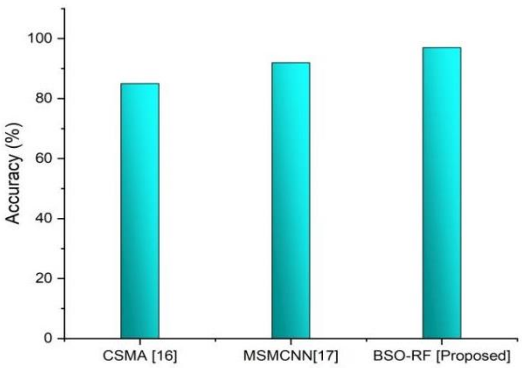


**Fig 1:** Result of accuracy

The algorithm's positive predictions are accurate to a certain degree, measured by precision. The ratio of true positives to the total of true positives and false positives is used to compute it. It shows how well the algorithm can prevent false positives. The precision outcome is shown in Figure 2. Comparatively, our suggested method (96%) outperforms the currently used methods (CMSA 83% and MSMCNN 90%).

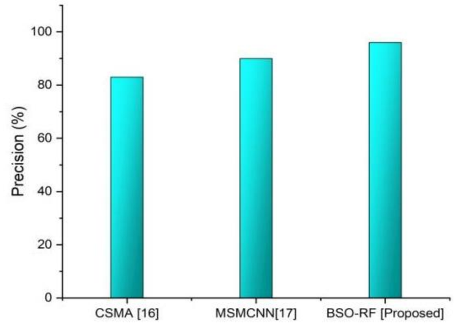


**Fig 2:** Result of prediction

The DSC calculates the degree of segmented data overlap areas and the actual data. This ratio is determined as the product of the volume of the segmented region and the ground truth multiplied by the point of intersection of the segmented region and the ground truth. The DSC results are shown in Figure 3. In terms of comparison, it demonstrates that the approach we propose (proposed method 93%) outperforms the currently used methods (CMSA 81% and MSMCNN 86%).

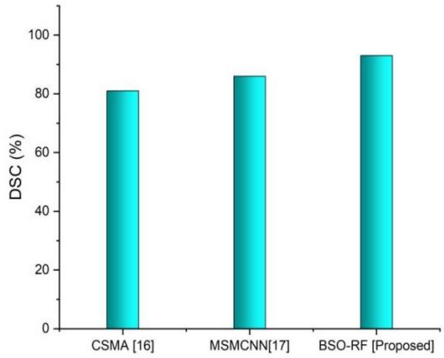


**Fig 3:** Result of DSC

The Jaccard index (JI), commonly called the Intersection over Union (IoU), quantifies how similar the segmented region and the real world are, just like the DSC. The intersection of the segmented area and the ground truth is divided by their union to arrive at the calculation. Figure 4 shows the results of the JI. It demonstrates that our suggested method (proposed method 0.8) outperforms the already used methods (CMSA 0.5 and MSMCNN 0.6).

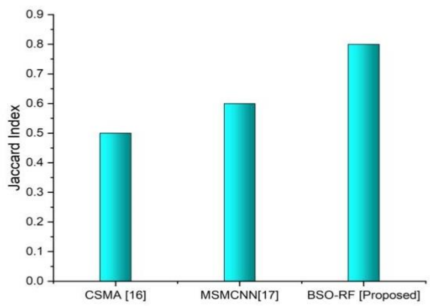


**Fig 4:** Result of JI

## 5. Conclusion

According to the study, the suggested BSORF method works better than other strategies at accurately segmenting multimodal medical images. The recommended method shows increased performance in recognizing and isolating important structures in complicated multimodal pictures by utilizing the BSO algorithm for feature selection and integrating it with the RF method. The studies' outcomes on several multimodal medical datasets show that the BSORF approach is effective and potentially has several therapeutic uses in medical imaging [18]. The research must discuss the proposed method's time complexity and computational needs. The BSORF approach's computational cost must be considered, especially when working with huge datasets or real-time applications. It would be beneficial to conduct user studies or work with medical experts to evaluate the clinical relevance and utility of the suggested method in particular therapeutic applications. This could lead to improvements.

## References


[1] Prakash, O., Park, C.M., Khare, A., Jeon, M. and Gwak, J., 2019. Multiscale fusion of multimodal medical images using lifting scheme-based biorthogonal wavelet transform. Optik, 182, pp.995-1014.

[2] Muhammad, G., Alshehri, F., Karray, F., El Saddik, A., Alsulaiman, M. and Falk, T.H., 2021. A comprehensive survey on multimodal medical signals fusion for smart healthcare systems. Information Fusion, 76, pp.355-375.

[3] Li, S., Zhang, C. and He, X., 2020. Shape-aware semi-supervised 3D semantic segmentation for medical images. In Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part I 23 (pp. 552-561). Springer International Publishing.

[4] Seo, H., Badiei Khuzani, M., Vasudevan, V., Huang, C., Ren, H., Xiao, R., Jia, X. and Xing, L., 2020. Machine learning techniques for biomedical image segmentation: an overview of technical aspects and introduction to state-of-art applications. Medical physics, 47(5), pp.e148-e167.

[5] Fang, L., Wang, X. and Wang, L., 2020. Multimodal medical image segmentation based on vector-valued active contour models. *Information sciences, 513*, pp.504-518.

[6] Shyni, H.M. and Chitra, E., 2022. A comparative study of X-ray and CT images in COVID-19 detection using image processing and deep learning techniques. Computer Methods and Programs in Biomedicine Update, p.100054.

[7] Huang, S.C., Pareek, A., Seyyedi, S., Banerjee, I. and Chest ren, M.P., 2020. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. NPJ digital medicine, 3(1), p.136.

[8] Shuvo, M.B., Ahommed, R., Reza, S. and Hashem, M.M.A., 2021. CNL-UNet: a novel lightweight deep learning architecture for multimodal biomedical image segmentation with false output suppression. Biomedical Signal Processing and Control, 70, p.102959.

[9] Yadav, S.P. and Yadav, S., 2020. Image fusion using hybrid methods in multimodality medical images. Medical & Biological Engineering & Computing, 58, pp.669-687.

[10] Lundervold, A.S. and Lundervold, A., 2019. An overview of deep learning in medical imaging focusing on MRI. Zeitschrift für Medizinische Physik, 29(2), pp.102-127.

[11] Du, J., Li, W., Lu, K. and Xiao, B., 2016. An overview of multimodal medical image fusion. Neurocomputing, 215, pp.3-20.

[12] Dolz, J., Gopinath, K., Yuan, J., Lombaert, H., Desrosiers, C. and Ayed, I.B., 2018. HyperDense-Net: a hyper-densely connected CNN for multimodal image segmentation. IEEE Transactions on medical imaging, 38(5), pp.1116-1126.

[13] Abumalloh, R.A., Nilashi, M., Ismail, M.Y., Alhargan, A., Alghamdi, A., Alzahrani, A.O., Saraireh, L., Osman, R. and Asadi, S., 2022. Medical image processing and COVID-19: a literature review and bibliometric analysis. Journal of Infection and public health, 15(1), pp.75-93.

[14] Choudhary, A., Tong, L., Zhu, Y. and Wang, M.D., 2020. Advancing medical imaging informatics by deep learning-based domain adaptation. Yearbook of medical informatics, 29(01), pp.129-138.

[15] Wang, R., Lei, T., Cui, R., Zhang, B., Meng, H. and Nandi, A.K., 2022. Medical image segmentation using deep learning: A survey. IET Image Processing, 16(5), pp.1243-1267.

[16] Ye, L., Rochan, M., Liu, Z. and Wang, Y., 2019. Cross-modal self-attention network for referring image segmentation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10502-10511).

[17] Rajasree, R., Columbus, C.C. and Shilaja, C., 2021. Multiscale-based multimodal image classification of brain tumor using deep learning method. Neural

Computing and Applications, 33, pp.5543-5553.

[18] Dr. Gupta, S. K. & Artono, B. (2022). Bioengineering in the Development of Artificial Hips, Knees, and other joints. Ultrasound, MRI, and other Medical Imaging Techniques. Technoarete Transactions on Industrial Robotics and Automation Systems (TTIRAS). 2(2), 10–15.

[19] Sunil Gavaskar, J., & Swetha, S. (2022). Multimodal Transport Solutions for Smart Cities. Technoarete Transactions on Advances in Computer Applications (TTACA), 1 (1), 15-19. doi: doi.org/10.36647/TTACA/01.01.A004