AN IIB-OPTIMIZED DEEP RESIDUAL NETWORK WITH GAN-BASED SEGMENTATION FOR AUTOMATED BRAIN TUMOR DETECTION FROM MRI IMAGES

INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE AND SUSTAINABLE TECHNOLOGIES
VOL 1 (01), JULY-DECEMBER 2026

Kalyani Chaunwal
Department of Computer Science and Engineering
JSS University, Noida, Uttar Pradesh, India
Shivangi Srivastava
Department of Computer Science and Engineering
JSS University, Noida, Uttar Pradesh, India
Vimal Gupta
Department of CSE-AIML
JSS University, Noida, Uttar Pradesh, India
Nitima Malsa
Department of Computer Science and Engineering
JSS University, Noida, Uttar Pradesh, India

ABSTRACT

Image processing of medical data relies heavily on the segmentation stage to distinguish and analyze relevant regions. Comparing photos to assess segmentation quality is an important component of tracking development in this field. Metric selection, the diverse conceptualisations of certain metrics across the literature, inefficiencies in the implementation of evaluation metrics posing challenges to large-scale data-processing, and a limited validation for fuzzy segmentation by the established metrics are some of the complications in analysing medical segmentation.

Keywords: *Evaluation metrics, Evaluation tool, Medical volume segmentation, Metric selection*

1. INTRODUCTION

Within clinical image analysis, 3D image segmentation serves as a fundamental stage in the processing of medical images. Segmentation methods improved precision (including reproducibility) and lower bias are among the key objectives of surgical planning, as they directly affect outcomes, such as the detection and tracking of tumor progression [1–3]. Warfield et al. [4] highlighted the clinical need for enhanced characterization of white-matter variations within brain tissue and showed that particular changes in white-matter regions are associated with certain brain diseases. Identifying shift patterns precisely is critical for timely disease detection and effective clinical surveillance. As a result, evaluating the precision and consistency of segmentation algorithms remains critical.

A 3D grid characterizes 3D diagnostic images, whose dimensions vary according to the anatomical regions being scanned and the imaging resolution.

The grid dimensions (w, h, d) represent the three dimensions of the image, namely the width, height, and depth respectively. A voxel is a three-dimensional point on the grid. A binary segmentation can be viewed as a partitioning scheme that determines whether individual voxels belong to a particular anatomical structure. Structures such as white matter, gray matter, brain lesions, and organs are examples of anatomical features. Segmentation evaluation involves comparing two segmentation results, where one is the output being assessed and the other provides the reference for comparison.

Medical image segmentations can be frequently ambiguous, implying where individual voxels may assume different membership values. A membership grade within [0, 1]. For instance, If the desired segmentation is obtained by averaging multiple segmentations corresponding to the same structure provided by different annotators, this is known as averaging.

Segments can be regarded as probability of voxels being a member of some particular classes in this context. To produce binary forms suitable for crisp segmentations, a common approach for evaluating fuzzy segmentations is to apply a threshold to the probability values. Thresholding, on the other hand, is not always successful. Furthermore, determining the threshold remains a challenge because the

evaluation outcomes are dependent on it. This is the purpose for developing measures that can compare fuzzy segmentations without sacrificing information. It's worth noting that partial volume is another typical interpretation of fuzzy segmentation. Another way of interpreting fuzzy segmentation considers the case where the voxel value represents the fraction of that voxel belonging to the class being considered. This interpretation can also be done using the definitions of the fuzzy metrics, presented in this work.

Different sorts of segmentation errors may be established according to various quality-related characteristics of 3D medical image segmentation. The choice of metric depends on the dataset and segmentation objective, and may need to account for one or several of these error types. Shi et al. [5] considers segmentation errors of four types: added regions, excess background, internal holes, and gaps along the boundaries. These errors can be characterized by considering the number and size of segmented objects, boundary precision, and the occurrence of holes. The requirements for evaluating medical segmentation were categorized by Fenster et al. [6] into accuracy, which reflects the agreement between the resulting segmentation and the ground-truth segmentation, and precision, which reflects repeatability, and and efficiency, which is primarily associated with processing time. Under the first category, they identified two quality characteristics, specifically, the definition of the object boundary (contour) along with the segmented object's volume. If the segmented objects have a small size, alignment—which describes the overall location of the segmented object—can become a more significant quality factor than either its size or boundary shape. Another difficulty in developing metrics is determining metric sensitivities. Sensitivity to specific attributes may limit detection of specific faults or contribute to their over- or underestimation. Outliers (small, extraneous segmented regions occurring beyond the main object, class imbalance (the proportion of the segmented region relative to the background), amount of segmented objects, and other factors influence the responsiveness of evaluation metrics. Another form of sensitivity concerns their ability to handle agreements that arise by chance. This is related to the baseline of a metric, which ideally should equal zero when segmenting at random, showing no resemblance [7]. A standardized framework for evaluating medical image segmentation is required, covering both the selection of metrics and the definitions assigned to them. To demonstrate the relevance of this, the section "Multiple definitions of metrics in the literature" includes instances of metrics that have multiple definitions in the literature" includes instances of metrics that have multiple definitions in the literature, each of which leads to distinct values, but are all referred to by the same name. The TREC EVAL tool enables consistent and standardized evaluation in the text retrieval area, avoiding such misunderstanding and providing a standardized benchmark for comparing text retrieval methods. In the domain of clinical imaging, such a tool does not exist. Gerig et al. [8] proposed the Valmet tool for medical volume segmentation evaluation. The evaluation capabilities of the tool are limited to five metrics, with several commonly used measures such as information theoretical metrics, some statistical measures namely Mahalanobis distance, and chance corrected metric including Kappa and modified Rand index remain unsupported in Valmet.

Additionally, this tool does not enable fuzzy segmentation.

The Insight Toolkit (ITK) is an open-source framework developed for handling a broad spectrum of medical image-processing and analysis tasks. Although this approach utilizes distance-transform-based filtering [9], it has certain limitations. In particular, the ITK library does not provide a comprehensive set of metrics required for medical segmentation analysis.

Second, the computational efficiency of several metrics is dependent on distance-transform filtering, resulting in increased processing and memory requirements as the volume size grows. The use of a bounding cube has been proposed to alleviate this limitation [6, 10], by restricting computation to the smallest enclosing region that contains both segmentations while excluding background voxels. Nevertheless, this strategy introduces several concerns. Large or spatially separated segmented regions can result in an excessively large enclosing volume, while isolated outliers may further expand its dimensions. Since the size of this region can affect the resulting metric values, the evaluation may become

sensitive to spatial characteristics that are not directly indicative of segmentation quality. Furthermore, as discussed in Section 1, comparison across multiple grid resolutions provides a more comprehensive assessment than evaluation at a single resolution [20]. This becomes particularly challenging for large medical volumes, including whole-body scans, where the associated computational and memory demands can be considerable.

The following are the contributions made by this paper:

- It gives an overview of 20 volume segmentation evaluation measures that were chosen after reviewing existing literature. In several instances, diverging definitions of metrics have been used in prior research, and standardized definitions are established.
- It incorporates metric-calculation algorithms designed to handle use of huge 3D image segmentations' nature as dense voxel distributions to operate best with them. Due to the rising extent of segmenting regions, such as segmentation of whole-body volumetric data, efficiency is becoming increasingly crucial.
- This work gives fuzzy definitions for all specified measures, allowing for the evaluation to account for ambiguity in medical picture segmentation.
- It offers generalized metrics for segmentation with numerous labels.
- In frequent scenarios of medical picture segmentation, it provides an Open-source implementation optimized for computational efficiency of all 20 metrics that beats advanced techniques.

2. PROPOSED IMPROVED INVASIVE BAT ALGORITHM BASED DEEP RESIDUAL NETWORK FOR BRAIN TUMOR DETECTION WITH MR IMAGES

The brain serves as the primary organ of the nervous system, which is responsible for controlling the body with the help of related parts. A brain tumor is characterized by the abnormal growth of cells within the brain. Consequently, detecting tumors at an early stage is important for successful treatment. In this work, an MR-based Deep Residual Network is employed to identify brain cancers at an early stage. Image preprocessing, segmentation, feature representation, and data augmentation, and tumor identification are all phases of the developed method. The image undergoes preprocessing to suppress unwanted noise and then input into the segmentation step, where GAN is used to segment the tumor. DFT, HOG, LOP, along with statistical descriptors, constitute some of the features derived from the segmented output. Statistical characteristics, meanwhile, include measures such as mean, standard deviation, entropy, and kurtosis. The Deep Residual Network then classifies the tumor according to whether it is physiological or pathological. Figure 1 illustrates the workflow of the proposed methodology.

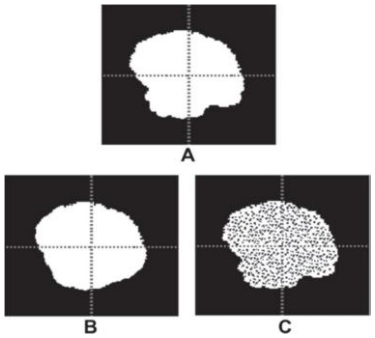

Figure 1 (The effect of segment density. Two segmentations b and c are compared with the corresponding ground truth (a). b has a solid structure while c has a lower density due to large number of tiny holes uniformly distributed inside it. Although c has a a higher accuracy of the boundary than b, all metrics, excepts MHD and HD, give b a higher score than (c))

Magnetic resonance imaging (MRI) is an established imaging technique that produces detailed, high-resolution views of internal body structures. It can provide useful details regarding tumor dimensions, position, and morphology without subjecting patients to ionizing radiation used in techniques such as CT and PET. Because of its strong soft-tissue resolution, MR imaging is commonly employed in the assessment of brain tumors. There is a compromise between resolution and signal to-noise ratio (SNR) while acquiring MR images, which degrades image quality. Because of the image noise and poor contrast in the MRI data, normal and malignant brain tissues are not readily distinguishable. As a result, image pre-processing is essential for noise suppression and improve contrast resolution across the segments. This process removes any image noise and reduces the effects of edge blurring. However following preprocessing, the resulting image is denoted by P and subsequently supplied to the segmentation stage. Automatic brain tumor segmentation using GAN faces a number of obstacles and issues. The task of segmenting the tumor is difficult since it combines MR image pathology, intensity, and shape analysis. Computer-assisted tumor segmentation remains challenging because the tumor often has indistinct boundaries and a shape that differs from the surrounding tissue. The segmentation approach is employed to separate the regions of interest from the background of an MR picture. Image segmentation helps lower the computational time required for processing. Image segmentation divides an image into connected spatial regions, with the pixels within each region sharing similar characteristics according to defined criteria. Brain tumor segmentation is subsequently performed using a GAN [18]. A conventional GAN consists of two components: the generator (R) and discriminator (D) (T). The generator produces a sample from noisy input data (m), whereas the discriminator examines the resulting sample and determines whether it is generated or authentic. The generator is trained to produce increasingly realistic samples, while the discriminator learns to differentiate generated samples from genuine ones. The generator therefore attempts to produce outputs that the discriminator will classify as genuine.

Consequently, the discriminator distinguishes generated samples from those originating from the target distribution. The segmented tumour outcome, on the other hand, is represented as accordingly. GAN's structure is depicted in Figure 2.

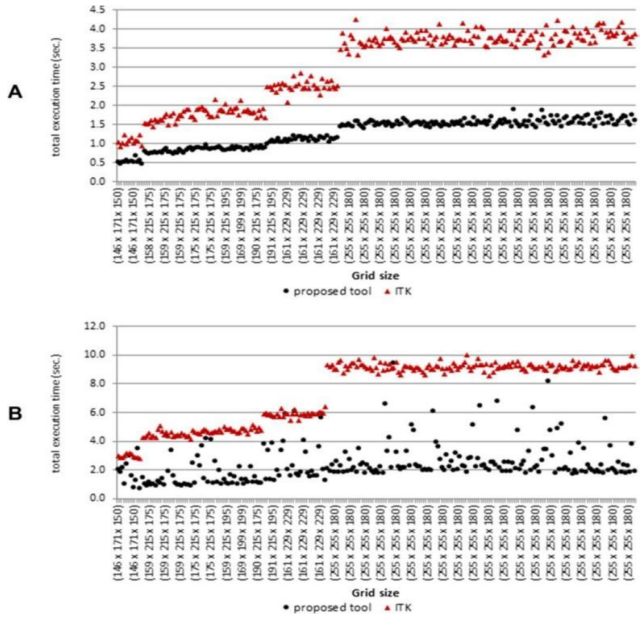

Figure 2 (Testing the proposed tool against the ITK implementation using brain tumor segmentation. Comparison between the performance of the proposed evaluation tool and the ITK Library implementation in validating 240 brain tumor segmentations against the corresponding ground truth using the HD in (a) and the AVD in (b). The grid size (width × height × depth) is on the horizontal axis and the run time in seconds is on the verticalaxis. The data points are sorted according to the total number of voxels, i.e. whd)

Following feature extraction from the segmented output, data augmentation can be performed more efficiently to produce superior tumor detection findings. To improve the resilience and For performance of classifiers, data augmentation refers to the generation of additional samples through modifications to the available training data. To expand the feature representation, every feature obtained from the segmented output is subjected to data augmentation. Consequently, the features obtained from the segmented output are handled individually by the data augmentation module, allowing multiple augmented representations to be produced. The resulting augmented data have a dimension of [0,000X50] and are produced through oversampling, with the minimum and maximum feature indices within the training set serving as the boundary thresholds when the DFT feature of size [1X50] is supplied to the data augmentation module. Applying the same oversampling procedure to the LOO feature of size [1X50] produces an augmented representation with the dimension [50,000X50].

Similarly, the enhanced output has a dimension of [50,000X50] when generated from the HOG feature of size [1X50]. Furthermore, statistical characteristics where each statistical descriptor—mean, standard deviation, entropy, and kurtosis—has a [1X1] dimension and produces an augmented output of [50,000X1]. The resulting output generated during tumor diagnosis, on the other hand, is stated as [50,000X154], which is inferred by the network in order to establish the tumor detection methodology. An IIB-based Deep Residual Network was employed for identifying brain tumors. Brain tumors represent a serious and potentially life-threatening form of cancer affecting both children and adults. Consequently, early detection is crucial for timely diagnosis and effective treatment. Brain tumor segmentation and detection are challenging tasks due to the variety of tumor appearance, shape, and size. Therefore, a deep learning classifier was developed to aid the diagnosis of brain cancers using augmented medical picture results. The technique of tumor detection in this study is done by using a Deep Residual Network trained through the IIB algorithm and by incorporating the augmented data, the deep learning model effectively processes MRI images, thereby improving its detection performance with higher reliability and robustness. It decreased training time while increasing training accuracy. As a result, the developed deep-learning model effectively resolves the computational complexity.

Deep Residual Network (ResNet) [6] is a neural network architecture designed to learn intricate patterns using multiple layers. It incorporates residual blocks together with convolution and pooling layers, which allows additional network depth while maintaining efficient information flow. Convolutional layer: The layer utilizes a kernel-based receptive field with an appropriate depth to capture the the features that are most informative within the augmented input. In general two-dimensional CNNs involve a smaller number of trainable parameters and provide greater weight sharing, while benefiting from localized receptive fields. Unlike fully connected layers, CNNs process the augmented input through localized connections and multiple filters to extract relevant features. During convolution, the filter is systematically applied across the input matrix, with element-wise multiplication followed by summation producing the corresponding output values. This sequence of mathematical operations forms the basis of convolutional layer computation. For generating the feature map output, this layer employs 64 kernels, each having dimensions of [3X3]. Pooling layer: Positioned between convolutional layers, this operation reduces the spatial dimensions of the feature map. It works well for controlling overfitting. The average pooling layer, along with the depth of the feature map, are determined so that each individual slice can be processed separately using an averaging operation in place of the receptive field. bin = bin cb h + 1 (15) bin = bin cb h + 1 (16) bin = bin cb h + 1 (17) where the input width, whereas bout and gout represent the width and height of the resulting output. In addition, cb represents the kernel width, whereas cg indicates its height. The extracted features are subsequently passed through a nonlinear activation function, allowing the network to capture more complex and nonlinear relationships.

The Rectified Linear Unit (ReLU) serves as the activation function in the proposed detection approach. It helps address the vanishing-gradient problem while also accelerating

model convergence. The ReLU operation can be expressed as ReLU = (0 B 0 B B 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0, The training dataset is divided by the deep learning model into smaller subsets, referred to as mini-batches, which are processed sequentially to balance computational cost with convergence.

Shortcut pathways between the input and output are introduced through residual blocks. Residual block-1 comprises 64 blocks with a [7x7] size, while residual blocks 2 and 3 each contain 128 blocks of [3x3] size. In addition, residual block-5 consists of 256 [3x3] blocks, whereas residual blocks 6 and 7 each comprise 512 blocks of [3x3] size. When the input and output dimensions are identical, the input can be linked directly to the output. If their dimensions differ, a dimension-matching parameter is applied to align the input with the output. d = X (B) + B (19) d = X (B) + CBB (20), where B and d denote the input and output of the residual blocks, respectively, while CB denotes the matching factor. The linear classifier first produces scores for the individual classes, after which the softmax function converts these scores into probability distributions. Connections between neurons across successive layers are established by the fully connected (FC) layer, ultimately producing the classification output. The input vector is normalized into a probability distribution using the softmax function. d = C.B + A (21) softmax (B) = e B to c=1 e Bc (22) Here, C denotes the weight matrix, A represents the bias, and T0 indicates the output dimension. IIB algorithm's algorithmic technique The suggested IIB algorithm's training process is built by combining IIWO[15] and BA[23]. IIWO is a The approach is a population-based optimization method inspired by the collective behavior of weed colonies, incorporating principles of chaos theory to guide the optimization process. Chaos theory, on the other hand, solves the optimization problem using a stochastic model. The non-repetitive nature of the approach promotes efficient global exploration by maintaining greater diversity within the population. This diversity enables the population to cover a broader region of the search space, thereby improving the exploration capability of the algorithm. The logistic function is a popular chaotic map. The optimal weed is selected to move the remaining weeds toward the most favorable location. The BA algorithm is a metaheuristic optimization method inspired by the echolocation mechanism of bats. Echolocation is used by all bats to detect and differentiate between potential food or prey and surrounding obstacles.

The bats move randomly through the solution space, with their velocity at a given point governed by the selected frequency and wavelength factors. The wavelength of the emitted pulses is adjusted dynamically, while the pulse emission rate varies according to the distance from the target. A solution is encoded as a vector, providing a structured representation that facilitates precise computation of the optimum. The solution is represented as M = [1X], where X denotes the weight obtained through the optimization process. The proposed IIB algorithm proceeds according to the following sequence of steps: Initialization: Let M denote the weed population within the solution space, while Mbest represents the best weed position.

Fitness calculation: the fitness value is determined by measuring the difference between the estimated and target values in such a the fitness function is defined as follows, where F represents the fitness measure, O corresponds to the original value, and d represents the detected result.

Overcrowding and parking are two of the most serious issues that large cities face. As a result, Automated Parking System Management is an effective strategy for managing parking spaces that makes use of web technology and databases. Smart parking is an electronic tool that allows users to use information technology to locate available parking spaces. Appropriate entries One of the most widely utilized types in smart

3. IMPLEMENTATION

The open-source project incorporates the 20 metrics obtained from the literature review and specified in [3]. The project is organized into several sections, including an 'Architecture' section, which describes its overall structure, and a 'Compatibility' section, which outlines the compatibility of the tools with different platforms. The section "Programming environment" contains information on the programming language, development framework, and execution environment. The ‘Efficiency optimization’ section includes several implementation specifics on the tool's

optimizations. Finally, under Section "Usage," several examples of usage are presented.

Architecture Evaluate Segmentation is a command-line utility designed to analyze two-dimensional or three-dimensional medical segmentations using the 20 evaluation metrics. Because it does not depend on a graphical user interface, the program can be incorporated into automated scripts and used to process a large number of segmentation results. The implementation is designed to exploit relationships between the different metrics. Common operations can be reused instead of being recalculated independently for every metric. This reduces unnecessary computation while reducing both processing time and memory consumption. The resulting evaluation data are presented in an easily interpretable format. They can also be exported as an XML file so that other software applications can process the results.

Compatibility

The program uses the **ITK library** for reading medical images. Because ITK supports a broad range of medical-image formats, Evaluate Segmentation can work with many different image formats without requiring separate implementations for each format. Another advantage is that when ITK introduces support for additional formats, the evaluation tool can benefit from that support without major modifications to its own image-reading functionality.

Environment for programming

Evaluate Segmentation is implemented using **C++** and the **CMake** build system. CMake allows applications written in native programming languages to be compiled across different operating systems and compiler environments. The project's source code and builds for different platforms are provided as open-source resources at https://github.com/codelab/Evaluate segmentation.

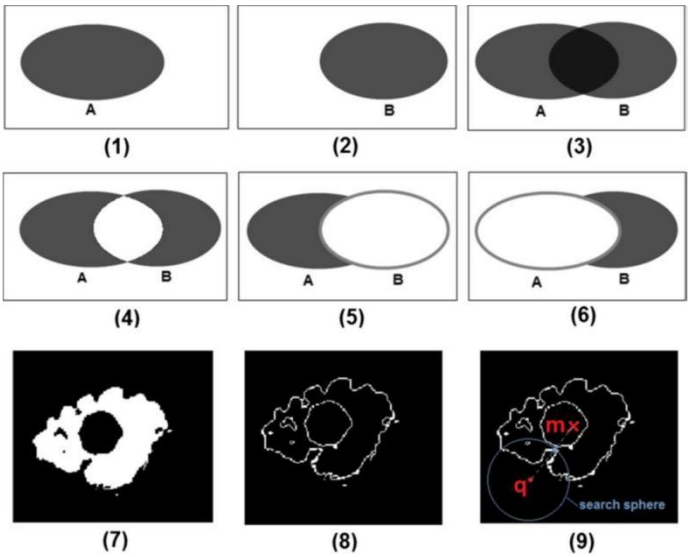

Figure 3. (Illustration of the optimizations used in calculating the average distance (AVD). In 1 and 2, the images A and B, defined on the same grid, are to be compared using the AVD. In 3, the intersection of the images is identified. In 4, the pairwise distance between points in the intersection is zero, therefore these distances are excluded from the calculation. In 5, to find the minimum distance from a point in A to the the image B, only the boundary voxels of B are considered. In 6, likewise to find the minimum distance from a point in B to the A, only the boundary voxels of A considered. In 7 and 8, the boundary voxels of a real segmentation of the edema of a brain tumor. In 9, to reduce the search space when searching the nearest neighbor, a search sphere with radius r is found by moving from the query q toward the mean m and considering the first point crossed on the boundary)

Increasing efficiency

Speed and memory consumption are important considerations when evaluating large 3D medical images. Whole-body medical scans can contain more than 100 million voxels, making inefficient pairwise calculations extremely expensive. The proposed implementation therefore uses optimization techniques to reduce unnecessary calculations and memory requirements. As image-grid dimensions increase, metrics that calculate pairwise distances between every voxel become increasingly expensive. Distance-transform-based approaches can therefore require significant processing time and memory. To address this problem, Evaluate Segmentation avoids directly relying on distance transforms for its calculations. Instead, it applies optimization strategies that improve computational efficiency while reducing memory consumption. For large images containing high-precision voxel values, a streaming mechanism provided by ITK is used to load and store the images efficiently. The representation can then use an appropriate data type for fuzzy voxel values, reducing memory requirements. The system also uses indexing techniques to represent image data efficiently. The **Hausdorff Distance (HD)** and **Average Hausdorff Distance (AVD)** require distance calculations between points belonging to the compared segmentations. Directly calculating every possible pair of distances can become computationally expensive for large images.To improve the efficiency of HD calculation, the proposed implementation uses **randomization and early-breaking techniques**. These methods identify unnecessary calculations and prevent the algorithm from examining voxel pairs that cannot influence the final result. The same optimization cannot be applied directly to AVD because AVD requires averaging distance information. Therefore, a modified nearest-neighbor method is employed. A three-dimensional cell grid is constructed around the point cloud, and a smaller search region is identified for each query point. This limits the number of candidate points that need to be examined and consequently reduces the number of distance calculations. Because medical segmentation regions are generally dense, the algorithm can exploit this characteristic to further improve nearest-neighbor searches. For example, when calculating distances from segmentation **A** to segmentation **B**, points located in the intersection of A and B can be ignored because their distance is zero. Similarly, instead of examining every point in B, the method can concentrate on boundary points because the nearest point encountered when moving from an interior point toward another segmentation will lie on its surface. A search radius can then be determined around the query point to restrict the search area. If no suitable point is found within that region, an exhaustive search is performed as a fallback.

Usage

Evaluate Segmentation is operated through the command line. The user supplies the reference segmentation and the segmentation under assessment. Optional parameters can specify which metrics should be calculated. If the -use option is not provided, the program calculates the available metrics by default. Individual metrics can also be selected by specifying their corresponding symbols. For metrics such as Hausdorff distance, additional parameters can be supplied to calculate a particular quantile rather than the complete distribution.

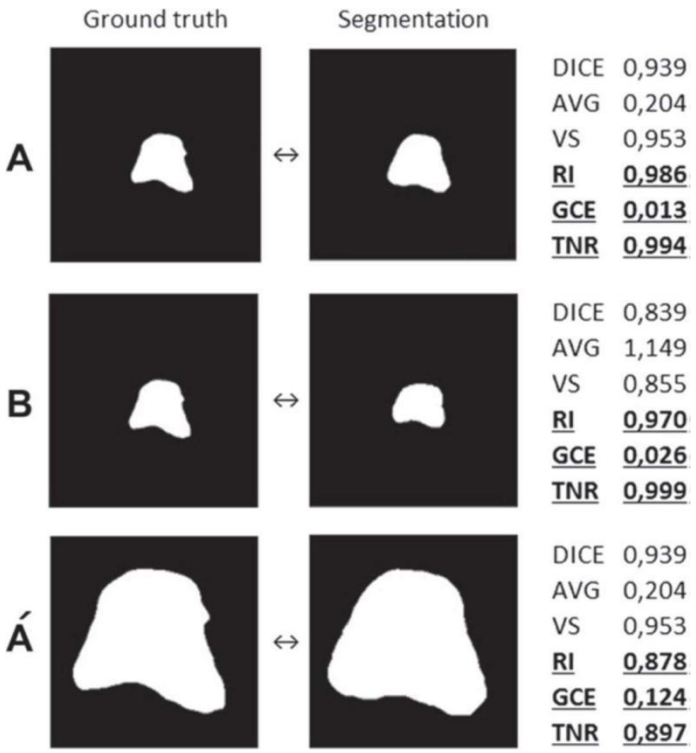

Figure 4. (The effect of decreasing the true negative (background) on the ranking. Each of the segmentations in A and B is compared with the same ground truth. All metrics assess that the segmentation in A is more similar to the ground truth than in B. In ´A, the segmentation and ground truth are the same as in A, but after reducing the true negatives by selecting a smaller bounding cube. The metrics RI, GCE, and TNR change their rankings as a result of reducing the true negatives. Note that some of the metrics are similarities and others are distances.

4. CONCLUSION

Using 20 assessment measures, we present a practical framework for assessing 3D medical image segmentations. The article focuses on evaluating the quality of medical image segmentation. The program aims to establish a standardized approach for assessing medical segmentation through a consistent collection of evaluation measures. The developed evaluation tool can be obtained fromhttp://github.com/codalab/EvaluateSegmentation. This tool's implementation uses strategies designed to address the challenges involved in medical segmentation assessment. The metrics are computed through algorithms selected and optimized to improve computational speed and efficiency.
Sufficient memory is necessary when processing images with large grid dimensions, including whole-body scans. Since evaluation measures differ in their characteristics, such as bias and sensitivity picking the right ones isn't easy.

The study examines the 20 implemented metrics, focusing on their qualities and suitability for evaluating segmentations based on specific needs and segmentations according to their specific characteristics. The analysis concludes by providing recommendations for selecting an appropriate subset. The usefulness of the 20 measures and provides recommendations for choosing suitable metrics according to the specific evaluation requirements. In this paper, In this study, a Deep Residual Network classifier is employed to develop a brain tumor detection approach. The deep learning classifier provides effective performance for both tumor segmentation and detection. During preprocessing, image imperfections are corrected and unwanted noise is suppressed. After that, the resulting image is passed to the segmentation stage, where the GAN technique partitions it to identify the relevant region more precisely. Features are subsequently derived from the segmented output, including LOOP, HOG, DFT, and statistical features. The data augmentation stage applies label-preserving

transformations to generate additional samples from the existing data. Finally, the Deep Residual Network determines whether the brain image corresponds to a normal or pathological case. The proposed IIB algorithm improves the training process of the deep learning classifier. Furthermore, evaluation on the BraTS dataset demonstrated the effectiveness of the proposed approach for brain tumor detection, with accuracy, sensitivity, and specificity values of 0.9256, 0.9003, and 0.9146, respectively. The Future research could investigate alternative deep learning classifiers for brain tumor segmentation and detection, with the potential to further improve tumor classification performance. An evaluation framework for 3D medical image segmentation was developed around 20 metrics identified from existing research. Implemented through the open-source *Evaluate Segmentation* project, the framework aims to provide a consistent basis for segmentation assessment while addressing the computational and memory demands associated with large volumetric images, including whole-body scans. Since segmentation metrics differ in their sensitivity, bias, and other properties, their suitability depends on the characteristics of the segmentation and the evaluation objective. The study consequently analyzes the behavior and applicability of the 20 measures and proposes guidelines for selecting the available metrics according to the desired segmentation characteristics and evaluation requirements. The developed framework uses a Deep Residual Network (ResNet) to detect brain tumors and is organized into a sequence of processing stages: preprocessing, segmentation, feature extraction, data augmentation, and classification. Initially, the input brain images undergo preprocessing to correct image imperfections and suppress unwanted noise. The resulting images are then passed to a GAN-based segmentation module, which partitions the images to accurately identify and isolate the relevant tumor regions. Following segmentation, discriminative features are extracted from the identified regions. The extracted feature set comprises Local Orientation Operator (LOOP), Histogram of Oriented Gradients (HOG), Discrete Fourier Transform (DFT), and statistical features, thereby capturing different characteristics of the segmented regions. The extracted samples are subsequently subjected to data augmentation using label-preserving transformations, generating new training samples while maintaining their original class labels. This increases the diversity of the training data and supports more effective learning by the classifier.

The augmented feature representations are then provided to the Deep Residual Network, which carries out the final classification of brain images as either normal or abnormal. The The proposed IIB algorithm is employed to optimize the training of the deep-learning classifier. Thus, the complete framework integrates GAN-based segmentation, handcrafted feature extraction, label-preserving data augmentation, and ResNet-based classification for brain tumor detection. Evaluation using the Fighshare dataset produced an accuracy of 0.9256, sensitivity of 0.9003, and specificity of 0.9146, demonstrating the framework's capability for brain tumor detection. Future investigations may also consider alternative deep-learning classifiers to further enhance tumor segmentation and classification performance.

REFERENCES

Zou KH, Warfield SK, Baharatha A, Tempany C, Kaus MR, Haker SJ, et al. Statistical validation of image segmentation quality based on a spatial overlap index. Academic Radiology.
2004;11:178–89.

Zou KH, Wells WM, Kikinis R, Warfield SK. Three validation metrics for automated probabilistic image segmentation of brain tumours. Stat Med.
2004;23:1259–82.

Kennedy DN, Makris N, Verne SC, Worth AJ. Neuroanatomical segmentation in MRI: Technological objectives. IJPRAI. 1997;11(8):
1161–87.

Warfield SK, Westin CF, Guttmann CRG, Albert MS, Jolesz FA, Kikinis R Fractional segmentation of white matter. In: Proceedings of Second International Conference on Medical Imaging Computing and Computer Assisted Interventions; 1999. p. 62–71. doi:10.1007/10704282_7.

Shi R, Ngan KN, Li S. The objective evaluation of image object segmentation quality. ACIVS. 2013;8192:470–9. doi:10.1007/978-3-319-02895-8_42.

Fenster A, Chiu B. Evaluation of segmentation algorithms for medical imaging. In: Conf Proc IEEE Eng Med Biol Soc. Shanghai; 2005. p. 7186–189.

Vinh NX, Epps J, Bailey J. Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. J Mach Learn Res. 2010;11:2837–854.

Gerig G, Jomier M, Chakos M. Valmet: A new validation tool for assessing and improving 3D object segmentation. In: Proceedings of the 4th International Conference on Medical Image Computing and Computer-Assisted Intervention; 2001. p. 516–23. doi:10.1007/3-540-45468-3_62.

Jr. Maurer CR, Qi R, Raghavan V. A linear time algorithm for computing exact Euclidean distance transforms of binary images in arbitrary dimensions. IEEE Trans Pattern Anal Mach Intell. 2003;25(2):265–70.

Udupa JK, LeBlanc VR, Zhuge Y, Imielinska C, Schmidt H, Currie LM, et al. A framework for evaluating image segmentation algorithms. ComputMed Imaging Graph. 2006;30(2):75–87.

Klement EP, Pap E, Mesiar R. Trends in logic. Netherlands: Springer; 2000. https://books.google.at/books?id=9HawzJbmXHUC.

Campello RJGB. A fuzzy extension of the Rand index and other related indexes for clustering and classification assessment. Pattern Recognit Lett. 2007;28(7):833–41.

Dice LR. Measures of the amount of ecologic association between species. Ecology. 1945;26(3):297–302.

Jaccard P. The distribution of the flora in the alpine zone. New Phytologist. 1912;11(2):37–50.

Cardenes R, de Luis-Garcia R, Bach-Cuadra M. A multidimensional segmentation evaluation for medical image data. Comput Methods Prog Biomed. 2009;96(2):108–24.

Al-Faris AQ, Ngah UK, Isa NAM, Shuaib IL. MRI breast skin-line segmentation and removal using integration method of level set active contour and morphological thinning algorithms. J Med Sci. 2013. doi:10.3923/jms.2012.286.291.

Crum WR, Camara O, Hill DLG. Generalized overlap measures forevaluation and validation in medical image analysis. IEEE Trans Med Imaging. 2006;25(11):1451–61.

Chinchor N. MUC-4 evaluation metrics. In: Proceedings of the 4th Conference on Message Understanding; 1992. p. 22–9. doi:10.3115/1072064.1072067.

Rijsbergen CJV. Information Retrieval, 2nd edn. Newton, MA, USA: Butterworth-Heinemann; 1979.

Martin DR, Fowlkes C, Tal D, Malik J. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proc. 8th Int'l Conf. Computer Vision. Vancouver, BC; 2001. p. 416–23.

Reddy AR, Prasad EV, Reddy LSS. Abnormality detection of brain mr image segmentation using iterative conditional mode algorithm. Int J Appl Inform Syst. 2013;5(2):56–66.

Vadaparthi N, Yarramalle S, Penumatsa SV, Murthy PSR. Segmentation of brain mr images based on finite skew gaussian mixture model with fuzzy c-means clustering and em algorithm. Int J Comput Appl. 2011;28(10): 18–26.
Reddy AR, Prasad EV, Reddy LSS. Comparative analysis of brain tumor detection using different segmentation techniques. Int J Comput Appl. 2013;82(14):14–28.

Brennan RL, Light RJ. Measuring agreement when two observers classify people into categories not defined in advance. Br J Math

Stat Psychol. 1974;27(2):. doi:10.1111/j.2044-8317.1974.tb00535.x.

Hallermeier E, Rifqi M, Henzgen S, Senge R. Comparing fuzzy partitions: A generalization of the rand index and related measures. IEEE T Fuzzy Syst. 2012. doi:10.1109/TFUZZ.2011.2179303.

Brouwer RK. Extending the Rand, adjusted Rand and Jaccard indices to fuzzy partitions. J Intell Inf Syst. 2009;32(3):213–35.

Anderson DT, Bezdek JC, Popescu M, Keller JM. Comparing fuzzy, probabilistic, and possibilistic partitions.
Trans Fuz Sys. 2010;18(5):906–18.

Rand WM. Objective criteria for the evaluation of clustering methods. J Am Stat Assoc. 1971;66(336):846–50.

Hubert L, Arabie P. Comparing partitions. J Classif. 1985;2:193–218.

Cover TM, Thomas JA. Elements of Information Theory. New York, NY,
USA: Wiley-Interscience; 1991.

Viola P, III Wells WM. Alignment by maximization of mutual information. international journal of computer. Int J Comput Vis. 1997;24(2):137–54.

Russakoff DB, Tomasi C, Rohlfing T, Jr Maurer CR. Image similarity using mutual information of regions. In: 8th European Conference on Computer Vision, ECCV; 2004. p. 596–607. doi:10.1007/978-3-54024672-5_47.

Meila M. Comparing clusterings by the variation of information. In Learning Theory and Kernel Machines. Berlin Heidelberg: Springer; 2003. p. 173–87.

Shrout P, Fleiss J. Intraclass correlations: Uses in assessing rater reliability. Psychol Bull. 1979;86(2):420–8.

Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960;20:37–46.

Hanley JA, Mcneil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143:29–36.

Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognit. 1997;30(7):1145–59.

Powers DMW. Evaluation: From precision, recall and Ffactor to ROC, informedness, markedness correlation. J Mach Learn Technol. 2011;2:37–63.

ACKNOWLEDGEMENT

This research was supported/partially supported by [JSSATE Noida]. We thank our colleagues from [JSS] who provided insight and expertise that greatly assisted the research, and lead to the outcome of the interpretations/conclusions of this paper.

Some potential thoughts are:

National Conference on Recent Innovations in Computing

28 September 2021 | JSSATE NOIDA INDIA