

Automated Brain Metastases Detection Framework for T1-Weighted Contrast-Enhanced 3D MRI

Engin Dikici¹, John L. Ryu², Mutlu Demirel³, Matthew Bigelow⁴, Richard D. White⁵, Wayne Slone, Barbaros Selnur Erdal⁶, and Luciano M. Prevedello

Abstract—Brain Metastases (BM) complicate 20–40% of cancer cases. BM lesions can present as punctate (1 mm) foci, requiring high-precision Magnetic Resonance Imaging (MRI) in order to prevent inadequate or delayed BM treatment. However, BM lesion detection remains challenging partly due to their structural similarities to normal structures (e.g., vasculature). We propose a BM-detection framework using a single-sequence gadolinium-enhanced T1-weighted 3D MRI dataset. The framework focuses on the detection of smaller (<15 mm) BM lesions and consists of: (1) candidate-selection stage, using Laplacian of Gaussian approach for highlighting parts of an MRI volume holding higher BM occurrence probabilities, and (2) detection stage that iteratively processes cropped region-of-interest volumes centered by candidates using a custom-built 3D convolutional neural network (“CropNet”). Data is augmented extensively during training via a pipeline consisting of random gamma correction and elastic deformation stages; the framework thereby maintains its invariance for a plausible range of BM shape and intensity representations. This approach is tested using five-fold cross-validation on 217 datasets from 158 patients, with training and testing groups randomized per patient to eliminate learning bias. The BM database included lesions with a mean diameter of ~5.4 mm and a mean volume of ~160 mm³. For 90% BM-detection sensitivity, the framework produced on average 9.12 false-positive BM detections per patient (standard deviation of 3.49); for 85% sensitivity, the average number of false-positives declined to 5.85. Comparative analysis showed that the framework produces comparable BM-detection accuracy with the state-of-art approaches validated for significantly larger lesions.

Index Terms—Magnetic resonance imaging, brain metastases, convolutional neural networks, deep learning, scale-space representations, computer-aided detection, medical image analysis.

I. INTRODUCTION

BRAIN metastases (BM) are disseminated cancer formations commonly originating from breast cancer, lung cancer, or malignant melanoma [1]. Detection of BM is a tedious and time-consuming manual process for radiologists, with no allowance for reduced accuracy; missed detections potentially compromise the success of treatment planning for the patient. Accordingly, computer-aided detection approaches have been proposed to assist radiologists by automatically segmenting and/or detecting BM in contrast-enhanced Magnetic Resonance Imaging (MRI) sequences, which is the key modality for the detection, characterization, and monitoring of BM. To this end, the most important imaging sequence is a T1-weighted image acquisition following intravenous administration of a gadolinium-based contrast agent. This sequence is particularly helpful for demonstrating vascularity within lesions as seen with BMs. Differentiating between tiny BM and vascular structure can be difficult, but in general, BMs are nodular, whereas vessels are tubular. Additional imaging, such as from T2-weighted or Fluid Attenuated Inversion Recovery (FLAIR) sequences, can be helpful to further characterize cysts and edema respectively. However, these features are more commonly seen with relatively larger lesions and contrast enhancement remains the optimal approach differentiating tiny BMs from benign lesions. This is especially true when 3D volumetric isotropic acquisitions are used (slices with thickness ≤ 1 mm), a key component in the detection of small brain lesions [2]. Different implementations of 3D T1-weighted images exist depending on the vendor; some examples include CUBE/BRAVO (from GE, SPACE/MPRAGE (from Siemens), and VISTA/3D TFE (from Philips).

Methods utilizing traditional image processing and machine learning techniques, such as template matching [3]–[5], 3D cross-correlation metrics [6], fuzzy logic [7], level sets [8], and selective enhancement filtering [9] are reported to produce promising results. In recent years, Convolutional Neural Network (CNN) [10] based approaches have started to be used extensively in a variety of medical imaging problems [11], [12], and this holds great promise for BM evaluation.

Manuscript received August 13, 2019; revised December 26, 2019 and March 12, 2020; accepted March 16, 2020. Date of publication March 23, 2020; date of current version October 5, 2020. (Corresponding author: Barbaros Selnur Erdal.)

Engin Dikici, Matthew Bigelow, Wayne Slone, and Luciano M. Prevedello are with the Department of Radiology, The Ohio State University College of Medicine, Columbus, OH 43210 USA (e-mail: engin.dikici@osumc.edu; matthew.bigelow@osumc.edu; wayne.slone@osumc.edu; luciano.prevedello@osumc.edu).

John L. Ryu is with the Department of Radiology, Perelman School of Medicine, The University of Pennsylvania, Philadelphia, PA 19104 USA (e-mail: john.ryu@pennmedicine.upenn.edu).

Mutlu Demirel, Richard D. White, and Barbaros Selnur Erdal are with the Department of Radiology, The Ohio State University College of Medicine, Columbus, OH 43210 USA, and also with the Department of Radiology, Mayo Clinic, FL 3224 USA (e-mail: demirel.mutlu@mayo.edu; white.richard1@mayo.edu; erdal.barbaros@mayo.edu).

Digital Object Identifier 10.1109/JBHI.2020.2982103

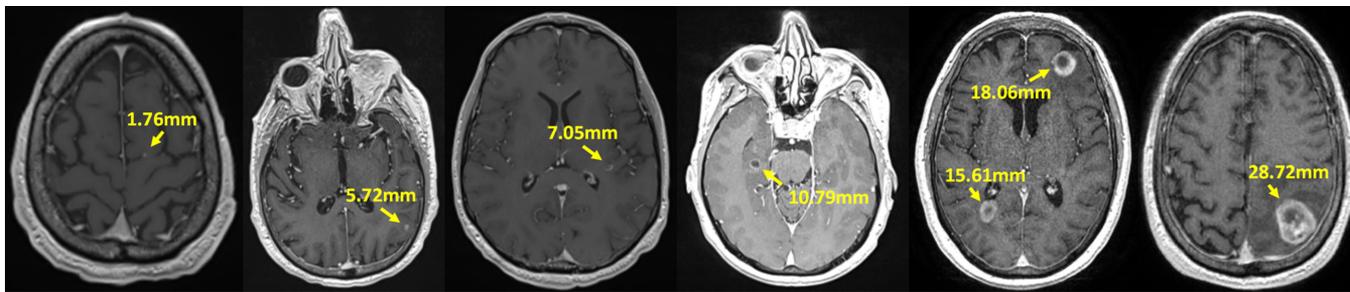


Fig. 1. Examples of brain metastases with their corresponding diameters. The study proposes a framework for detection of BM with diameters of ≤ 15 mm. The 2D axial view images are each scaled independently.

To our knowledge, the application of a Deep Neural Network (DNN) for segmentation of BM in MRI datasets was first introduced by Losch *et al.* [13]. Besides analyzing the impact of different network structures on the segmentation accuracy, their study also showed that a DNN can produce comparable or even better results with respect to previously reported state-of-art approaches. However, a limitation of their approach was a significant reduction in accuracy for the segmentation of tumors with sizes below 40 mm^3 .

Charron *et al.* [14] used DeepMedic neural network [15] for segmenting and detecting BM in multi-sequence MRI datasets as input, including post-contrast T1-weighted 3D, T2-weighted 2D fluid-attenuated inversion recovery, and T1-weighted 2D sequences. The study involved investigation of the impacts of epoch, segment, and/or batch sizes on overall accuracy, thus providing a well-documented hyper-parameter optimization process. The BM considered in their study had a mean volume of 2400 mm^3 , and the system detected 93% of lesions whereas producing 7.8 average false-positive detections per patient.

Liu *et al.* proposed a modified DeepMedic structure, “En-DeepMedic” [16], with the expectation of improved BM segmentation accuracy and higher computational efficiency. The approach was validated with both the BRATS database [17] and their post-contrast T1-weighted MRI collection of brain metastases with a mean tumor volume of 672 mm^3 . The system yielded an average Dice similarity coefficient of 0.67, where the detection false-positive rate in connection to the sensitivity percentage is not reported.

More recently, Grøvik *et al.* [18] demonstrated the usage of 2.5D fully CNN, based on GoogLeNet architecture [19], for detection and segmentation of BM. Their solution utilized multiple sequences of MRI for each patient: T1-weighted 3D fast spin-echo (CUBE), post-contrast T1-weighted 3D axial IR-prepped FSPGR, and 3D CUBE fluid-attenuated inversion recovery. Their database included 156 patients, with testing performed on 51 patients. For the detection of BM, at 83% sensitivity, average number of false-positives per patient is reported as 8.3.

The motivation for our study is to provide a BM-detection framework for 3D T1-weighted contrast-enhanced MRI datasets that focuses on small lesions (≤ 15 mm) with an average volume of only $\sim 160 \text{ mm}^3$ (see Fig. 1). Such tiny lesions are difficult for even experienced neuroradiologists to detect, and missed lesions can lead to inadequate or delayed treatment. To our knowledge, no prior work focused on BM with volumes

smaller than 500 mm^3 . Detection of small lesions is particularly important given the clinical challenge they represent and due to recent paradigm shift in how these lesions are treated with radiation. In the past, patients with multiple intracranial metastases were treated with whole brain radiation, making detection of individual lesions not as crucial. However, due to long-term cognitive decline associated with whole brain radiation, recent radiation treatment regimens target individual lesions, consequently making detection of even a tiny lesion crucial for the appropriate treatment [20].

This report first provides the following components of the detection framework: (1) Candidate BM selection procedure, (2) training strategy, (3) data augmentation pipeline, and (4) CNN architecture. Next, a description of the medical data used in the study, BM data statistics, and its acquisition are presented. The evaluation criteria are then defined, and the motivation behind the use of the average number of false lesion-detections per patient in connection to the detection sensitivity is justified. Next, the results are provided based on a five-fold Cross-Validation (CV) executed on 217 datasets. Finally, it is concluded with: (1) Comparisons with other state-of-art techniques in the field, (2) summary of the novelties of the introduced study, (3) system limitations, and (4) future work indicators.

II. METHODS AND MATERIALS

The BM-detection framework consists of two main components: (1) Candidate-selection step, and (2) a classification stage. First, the input MRI volume is processed using an information-theory based approach for detection of image points with high probability of representing BM. Next, volumetric regions centered by these candidate locations are iteratively fed into a custom-built CNN, *CropNet*, with extensive data augmentation, including rigid and non-rigid geometric transformations and intensity-based transformations. *CropNet* is a classification network, trained and validated to determine the probability of a given volumetric image to contain a BM. Algorithmic details of these stages are further described in the following subsections.

A. Metastasis Candidate Selection

The visual appearance of metastatic masses can be generalized to blob-shaped formations either with relatively brighter or darker interiors (i.e., due to central necrosis). Blob-detection has been previously addressed using various generalized scale-space methods [21], including the Laplacian of Gaussian (LoG)

approach [22]. In the proposed detection framework, LoG is utilized for detecting BM candidates for a given MRI volume as it: (1) Avoids image noise via its inherited Gaussian filtering properties, (2) holds few parameters to optimize, and (3) robustly detects BM candidates, with sensitivity reported in the Results section.

Yu *et al.* deployed LoG in the detection stage of their BM segmentation approach for MRI images [23], solidifying the applicability of LoG in the domain of our study. We further enhance the approach with sensitivity constraints and use it in candidate selection.

Given volumetric image data V , scale-space representation can be defined as,

$$L(x, y, z; s) = G(s) * V, \quad (1)$$

where s is the scale, and L gives the scale-space representation at s . Accordingly, the scale-normalized Laplacian operator is:

$$\nabla_{norm}^2 L = s(L_{xx} + L_{yy} + L_{zz}). \quad (2)$$

Local optima of the above equation, which are maxima/minima of $\nabla_{norm}^2 L$ with respect to both space and scale, represents the blob center positions [22].

The BM candidate-selection process aims to determine a set of image points that are closely located to the metastatic mass centers. Keeping the candidate list as short as possible is one of the main objectives for the process. However, the sensitivity of the framework needs to be maintained, which implies a comprehensive list of candidates. As these objectives are working against each other, the optimization process can be described as a minimax problem:

$$\arg \max_p (Sv(LoG(p, V), M)), \quad (3)$$

$$\arg \min_p (|LoG(p, V)|), \quad (4)$$

where Sv defines the sensitivity of the system based on (1) M representing the list of actual BM centers, and (2) $LoG(p, V)$ denoting candidate points selected for input volume V with LoG parameters of p . As the sensitivity of the system is the major criterion in this study, we propose a solution where the sensitivity portion of the equation is constrained as

$$\arg \max_{p, Sv \geq \theta} (S(LoG(p, V), M)), \quad (5)$$

with θ giving the minimal allowed sensitivity (e.g., 95 percent), and p is found via grid-search [24] constrained with Equation-4.

B. Network Training

The DNN described in the following subsection aims to classify each BM candidate as positive (implies that the candidate point holds high probability for being a center of metastatic mass) or negative. The proposed BM candidate selection method generates candidates in magnitudes of thousands (please refer to Results section for actual numbers), where only a few of these are true BMs. Thus, the network training should factor in highly unbalanced class representations. The proposed detection framework addresses this using (1) random paired data selection strategy, and (2) on the fly data augmentation stage aiming

to represent the covariance of tumor representations using a stochastic methodology.

During the training of the DNN, at each batch iteration, a pair of positive and negative samples are selected from each dataset randomly, producing a batch of $2N$ samples where N is the number of training cases. Next, the given batch is augmented on the fly [25], and the DNN is trained with the augmented batch (see Fig. 2, row A). The term ‘‘epoch’’ is not used in this definition; as in the proposed framework, the samples are processed in a random pair basis, whereas epoch commonly refers to complete pass through all training data.

The augmentation process is the key for the introduced detection framework’s learning invariance. The BM sample count is a small fraction of the total amount of samples—the learning process heavily depends on properly generalizing intensity and shape variations of BM. The importance of data augmentation for general computer vision and similar medical imaging scenarios are further described in [26] and [27], respectively. The detection framework deploys an augmentation pipeline consisting of random (1) elastic deformation, (2) gamma correction, (3) image flipping, and (4) rotation stages (see Fig. 2, row B). In the following subsections, technical details for random elastic deformation and random gamma correction augmentations are provided. Next, the CNN, which processes the augmented positive and negative sample volumes, is further described.

1) Random Elastic Deformations: The applicability of elastic deformations as a data augmentation step for detection of prostate cancer in multi-parametric MRI was illustrated by Yang *et al.* [28]. In their study, to augment a given 2D-MRI image, a random group of control points and their corresponding random 2D relocation vectors were first determined. A thin-plate transformation [29] for the given control point and relocation vector pairs was then computed to generate a 2D elastic deformation field. For a similar medical application, Le *et al.* [30] showed the advantages of using both rigid and non-rigid (i.e., elastic) deformations during data augmentation.

In our study, plausible non-rigid augmentations of the BM regions are produced by a fully 3D approach that does not require control points: the method generalizes the random elastic deformation field generation algorithm proposed by Simard *et al.* [31] to 3D. More explicitly, for a given volumetric image data V , random displacements fields ΔV_x , ΔV_y and ΔV_z are defined, where each of these has similar dimensions as V , and their voxels hold random values picked from a uniform distribution defined in the range of $[0, 1]$. Next, these random fields are smoothed with a zero-centered Gaussian kernel with a standard deviation of σ (defined in mm). Finally, the deformation field is scaled with an elasticity coefficient α . Choice of σ causes elastic deformation to be (1) pure random with $\sigma \leq 0.01$, and (2) almost affine with $\sigma \geq 5$, whereas α determines the magnitude of the introduced local deformations (Fig. 3).

The usage of elastic deformations in the augmentation stage is crucial for the proposed framework, as it facilitates the generation of a conceivable BM shape domain. However, the algorithm needs to be used with well-tested parameters to ensure the viability of the augmented BM samples. In their paper, Simard *et al.* suggest the usage of $\sigma = 4$ and $\alpha = 34$, as it yielded

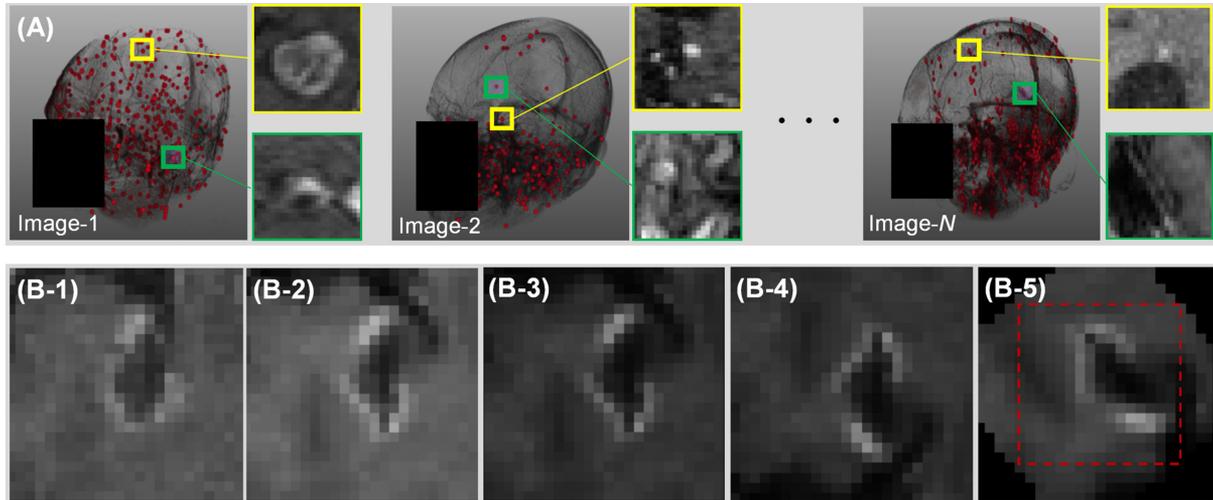


Fig. 2. Row A: Compilation of the positive & negative pair batch is represented. Positive and negative samples (shown with yellow and green rectangles respectively), are selected from BM candidates shown with red spheres in each dataset. Row B: Each positive sample goes through augmentation process: (B-1) mid-axial slice of an original cropped sample, (B-2) random elastic deformation is applied, (B-3) random gamma correction is applied, (B-4) sample volume is randomly flipped, and (B-5) sample volume is randomly rotated. The middle part of the randomly cropped volume, shown with a dashed red square in (B-5), is used for the training. Face regions are covered to protect patient privacy.

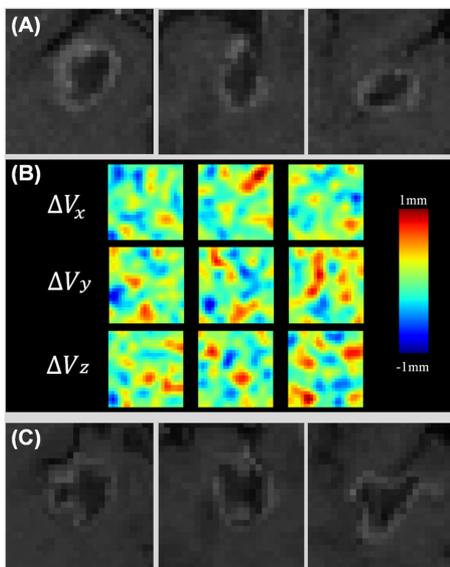


Fig. 3. The original cropped volume of a metastatic tumor mass (A), random displacement fields for x, y and z axes (B), and the corresponding deformed volume (C) are each shown from mid-axial, sagittal, and coronal views.

the best results in their analyses. Our framework adopted those optimal parameters after visual inspections by a medical expert.

2) Random Gamma Corrections: In MRI, tissues do not have consistent intensity ranges, such as in computed tomography. Usage of bias field correction might improve the predictability of tissue intensities. However, its success is limited due to machine-dependent parameters [32]. Medical image processing algorithms, both information-theory and DNN based, benefit from understanding the probabilistic distributions of tissue intensity values. One way to achieve this goal is the normalization of image intensities in MRI to represent the target tissues with

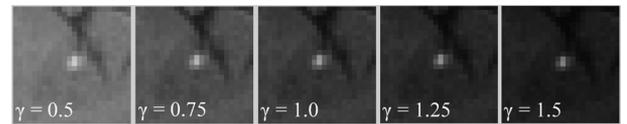


Fig. 4. The effects of gamma correction on region centering ~ 2.2 mm diameter metastasis (mid axial slice of a cropped 3D volume). Please note that $\gamma = 1.0$ represents the original image.

predefined intensity ranges [33]. Using even order derivatives of the histogram [34], Gaussian intensity normalization of selected tissues [35], and utilizing the median of intensity histogram [36] are some of the approaches introduced for that purpose. However, these methods are shown to be prone to errors as they aim to define approximations to non-linear intensity matching problems. The region-based approach [37], is shown to be effective, as it divides the spatial domain into smaller regions to address this limitation via piecewise linear approximations.

In the proposed framework, a novel form of the region-based strategy is introduced; random gamma corrections are applied to cropped volumetric regions during the augmentation stage [38]. Accordingly, the framework (1) does not make any assumptions about the histogram shape or intensity characteristics of given MRI datasets, and (2) avoids losing or corrupting potentially valuable intensity features, which is a common disadvantage of image intensity normalization-based methods.

Gamma correction of given volumetric data is given by,

$$V_G = V_N^{(1/\gamma)}, \quad (6)$$

Where V_N is the intensity scaled volumetric image data in $[0, 1]$ range, γ is the gamma value, and V_G is the gamma-corrected volumetric image data, which is also intensity scaled (see Fig. 4).

In the detection framework, the gamma correction augmentations are utilized by randomly picking γ values from a uniform

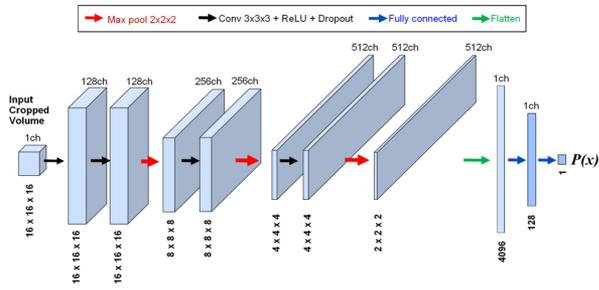


Fig. 5. CropNet-b2-16 mm: Input of this CNN is 16 mm x 16 mm x 16 mm isotropic region-of-interest, and each resolution level consists of two identical blocks, where the output is a scalar in range of $[-1, 1]$.

distribution defined in $[0.8, 1.2]$ range, determined empirically by investigating the visual appearance of gamma-corrected volumetric regions.

3) *Network Architecture*: The CNN introduced in this study (i.e., *CropNet*) has an input layer with an isotropic-sampled volumetric region of interest (ROI), where each voxel represents 1 mm^3 . Please note that the input volume's edge length is used in model naming, such as *CropNet-[c]mm*, where c is the volume's edge length in mm. Besides, the model follows a typical contracting path structure: Each resolution level is formed using stacked blocks each consisting of convolution, rectified linear activation unit (ReLU) and dropout layers. Block count per resolution level is another configurable parameter for the introduced network, hence, included in the naming convention as *CropNet-b[B]*, where B denotes the number of blocks per level. The network's downsampling is performed via $2 \times 2 \times 2$ max-pooling, followed with channel doubling. The output is a one-dimensional scalar produced via the sigmoid activation layer, which holds value in the range of $[0, 1]$ representing the likelihood of a given ROI to contain a metastatic mass. The network's convolution layers are initialized using Glorot uniform initializer as described in [39].

In Fig. 5, the formation of network architecture is illustrated for two blocks and 16 mm edge length (*CropNet-b2-16mm*), thus the reader can associate naming convention with the CNN. As described in the Results section, the study employs 16 mm version of CropNet, as (1) the target objects have diameters smaller than 15 mm, and (2) *CropNet-b2-16mm* produced comparable performance and allowed faster training with respect to its higher edge length versions (i.e., 32 and 64 mm).

4) *Data Preprocessing*: During the data preprocessing stage, all datasets are resampled to have $(1 \text{ mm} \times 1 \text{ mm} \times 1 \text{ mm})$ voxels, as CropNet requires isotropic sampled ROIs at its input layer. No further morphological, or intensity altering transformations are applied to the data during this stage.

III. DATABASE

A. Data Collection

This retrospective study was conducted under Institutional Review Board approval with a waiver of informed consent (institutional IRB ID: 2016H0084). A total of 217 post-gadolinium T1-weighted 3D MRI exams were collected from 158 patients:

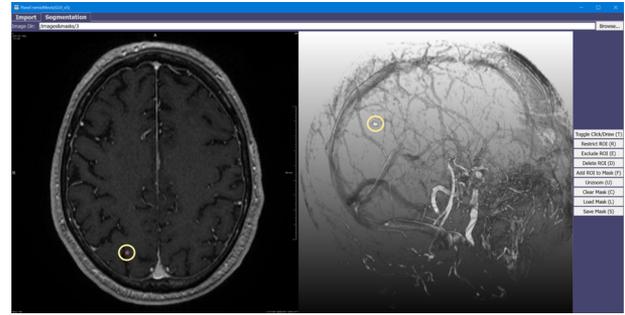


Fig. 6. The screenshot of the custom tool allowing medical experts to create, edit, save/load segmentation masks of BMs in MRI images. The tool provides 2D axial view, 3D view, and various manual editing tools.

113 patients with a single dataset, 33 patients with 2 datasets (i.e., one follow-up examination), 10 patients with 3 datasets, and 2 patients with 4 datasets. The images were collected from 8 scanners, where the acquisition parameters for each are summarized in Table I. Dotarem (gadoterate meglumine) was used as the contrast agent.

Two of the major study selection parameters were that (1) none of the datasets involved lesions with diameter of 15 mm or larger, and (2) motion degraded studies were included.

Ground-truth BM segmentation masks were prepared by a radiologist, using a custom-built tool for the project [40]. The tool was developed using MeVisLab 2.8 (medical image processing and visualization framework developed by MeVis Medical Solutions AG), and it allows users to load volumetric MRI datasets, manually delineate the borders of BM, and edit the existing segmentation masks if needed (see Fig. 6).

B. Brain Metastases

The database included 932 BMs where, (1) mean number of BMs per patient is 4.29 ($\sigma = 5.52$), median number per patient is 2, (2) mean BM diameter is 5.45 mm ($\sigma = 2.67 \text{ mm}$), median BM diameter is 4.57 mm, and (3) mean BM volume is 159.58 mm^3 ($\sigma = 275.53 \text{ mm}^3$), median BM volume is 50.40 mm^3 . Fig. 7(A, B and C), provides the histograms for each of these distributions.

For better understanding of the localization of BMs included in our database, all BMs are registered on a reference MRI image, and the probability density function is generated for multiple projections in Fig. 7(D). The volumetric registration for this illustration is performed by maximizing the mutual information between the reference MRI volume, and the rest of the volumes in the database iteratively, maximizing:

$$I(V_C, V_{Ref}) = H(V_{Ref}) - H(V_{Ref}|V_C), \quad (7)$$

where V_C is the floating volume (i.e., any volume picked from the database), V_{Ref} is the reference volume, $H(V_{Ref})$ is the Shannon entropy of the reference volume, and $H(V_{Ref}|V_C)$ is the conditional entropy. Rigid registration, optimizing translation and rotation parameters, is utilized in our visualization. The interested reader may refer to [41] for further details on mutual information's usage in medical image registration.

TABLE I
SCANNER PARAMETERS

Scanner	MF ^a (T)	TR ^b range (ms)	TE ^c range (ms)	Slice thickness range (mm)	Pixel size ^d range (mm)	Imaging frequency range (MHz)	Flip angle range (degrees)	Exam #
Siemens Aera ^e	1.5	[9.3, 9.7]	[4.4, 4.7]	1.0	[0.78, 0.97]	63.6	20	54
Siemens Avanto ^e	1.5	[9.7, 10]	[4.2, 4.8]	[0.9, 1.0]	[0.43, 0.86]	63.6	[15, 20]	17
Siemens Espree ^e	1.5	10	[4.5, 4.7]	1.0	[0.78, 1.0]	63.6	20	26
Siemens Skyra ^e	3.0	[6.2, 6.5]	2.46	[0.8, 0.9]	[0.65, 0.78]	123.2	[10.5, 12]	34
Siemens TrioTrim ^e	3.0	6.5	2.45	[0.8, 0.9]	[0.65, 0.73]	123.2	10.5	4
Siemens Verio ^e	3.0	[6.5, 9.0]	[2.4, 4.9]	[0.8, 0.9]	[0.65, 0.78]	123.2	10.5	28
GE Optima MR450w ^f	1.5	10.3	4.2	1.0	0.49	63.9	20	26
GE Signa HDxt ^f	1.5	[9.2, 10.3]	4.2	1.0	[0.49, 0.98]	63.9	20	28

^aMagnetic field strength, ^brepetition time, ^cecho time, ^dpixel size is same in x and y directions.

^eSiemens Healthcare, Erlangen, Germany.

^fGE Healthcare, Milwaukee, Wisconsin, USA.

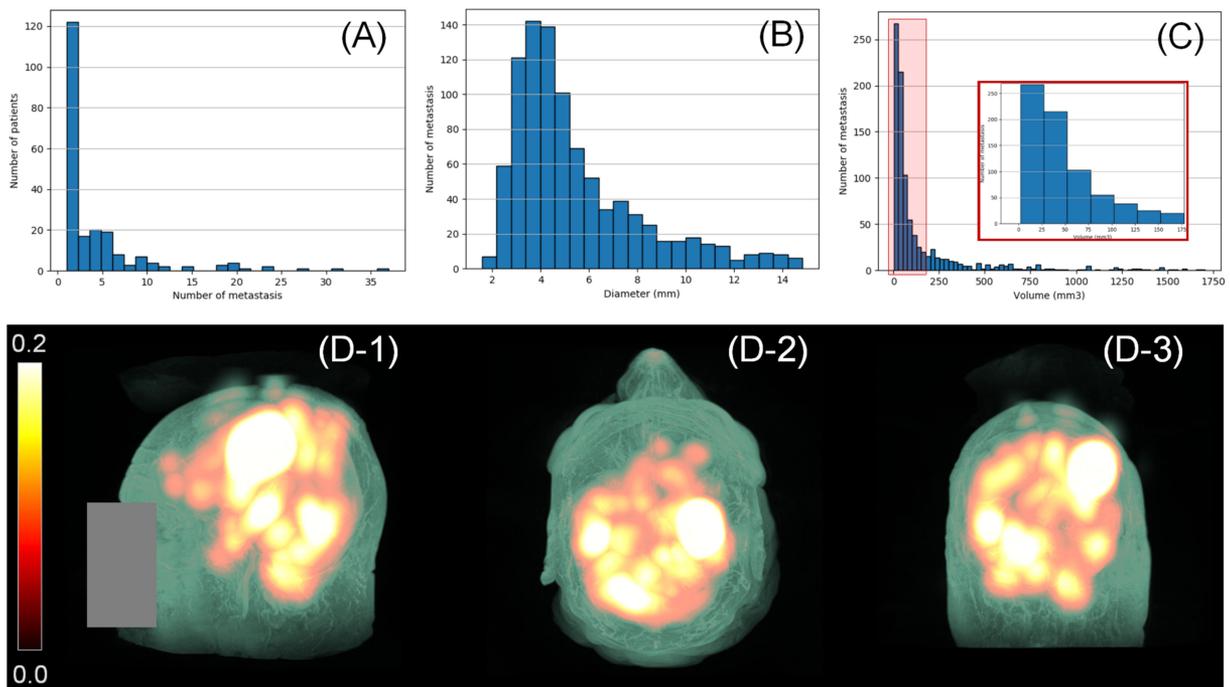


Fig. 7. The histograms for (A) number of BM per patient, (B) diameters of BM, and (C) volumes of lesions in BM database are shown. Below, the BM probability density function's projections on left sagittal (D-1), axial (D-2), and coronal (D-3) planes are provided. Face region is covered to protect patient privacy.

C. Evaluation Metric

The clinical applicability of a BM-detection algorithm was assessed by measuring (1) the sensitivity and (2) the average number of false lesion-detections for a given sensitivity.

As a screening tool, sensitivity of the system is expected to be high: In a typical deployment scenario of a detection algorithm, the appropriate operating point, maximizing the sensitivity whereas minimizing the average false lesion-detections per patient, needs to be adjusted by a medical expert. Therefore, we plot our performance metrics (i.e., sensitivity vs average number of false-positive detections per patient - AFP) at various output threshold settings (~ 0 - low likelihood and ~ 1 - high likelihood of metastasis). Accordingly, state-of-art approaches [13], [14], [18] follow a similar reporting methodology.

IV. RESULTS

The detection framework is validated using 5-fold CV. Folds are generated based on patient, which ensures each patient is located either in a training or testing group for each fold (e.g. datasets from Patient-A are all located either in training or testing group for fold-n) for eliminating the learning bias. Accordingly, the bins included datasets from 31, 31, 32, 32 and 32 patients, respectively. For each CV fold, four bins are used for the training and validation, and a single bin is used for the testing.

For the candidate selection stage of the framework, Laplacian of Gaussian parameters are optimized from the training bins with the constraint of setting minimal sensitivity to 95% (see Equation-5). These parameters include (1) minimal and maximal standard deviations for the Gaussian kernel, and (2) the absolute

TABLE II
OPTIMAL LOG PARAMETERS AND KERNEL RADIUS RANGE

CV fold	Min stdev (mm)	Max stdev (mm)	SSM (%) ^a	Radius Range (mm) ^b
1	1	4	1.0	[2,7]
2	1	4	0.5	[2,7]
3	1	5	1.0	[2,9]
4	1	4	1.0	[2,7]
5	1	5	1.5	[2,9]

^aSSM is described as the percentage of maximum image intensity.

^bGaussian kernel's radius range is derived from the minimal and maximal standard deviations.

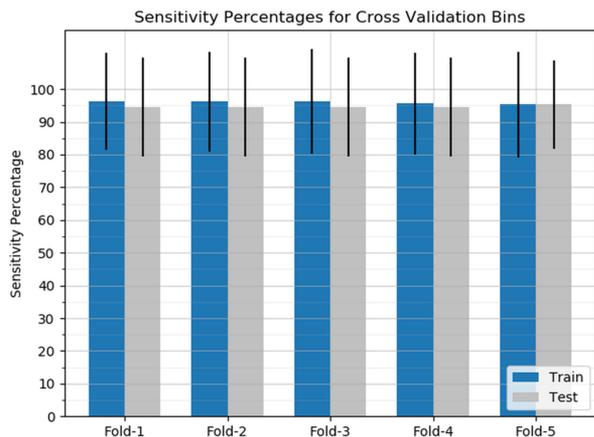


Fig. 8. Candidate selection procedure's sensitivity percentages for each fold's training (blue) and testing (silver) groups are represented. Sensitivity standard deviations are also shown with bold lines on each block.

lower bound for scale-space maxima (SSM), also referred to as LoG threshold in the literature [22]. During this optimization: (1) The minimal and maximal standard deviations were searched in the range of [1, 6] mm with the step size of 1 mm, and (2) SSM was searched in the range of [0.5, 2.5]% with the step size of 0.5%. In the utilized image processing library [42], the LoG method's Gaussian filter adapts its kernel radius based on the standard deviation; $kernel\ radius = \sqrt{3} \cdot stdev$. Table II summarizes the found optimal LoG parameters and the corresponding Gaussian kernel radius range for each CV fold. The candidate-selection procedure achieved (1) a mean sensitivity of 95.8, where the sensitivity for training and testing groups of each fold are represented in Fig. 8, and (2) produced 72623 candidates on average ($\sigma = 12518$) for each 3D dataset. Processing time for each dataset is ~ 30.6 seconds (using a 3.5 GHz Intel Core i7-5930K CPU).

The framework contained *CropNet*-b2-16 mm for processing the BM candidates and providing the final detection results. The network processed cubic ROIs with 16 mm edges and each resolution level included two blocks with layers as described in Section II. The dropout rate was set to 0.15 throughout the network. The optimization was performed using Adam algorithm [43], where the learning rate was 0.00005, and the exponential decay rates for the first and second moment estimates were set as 0.9 and 0.999 respectively. Binary cross-entropy was used as

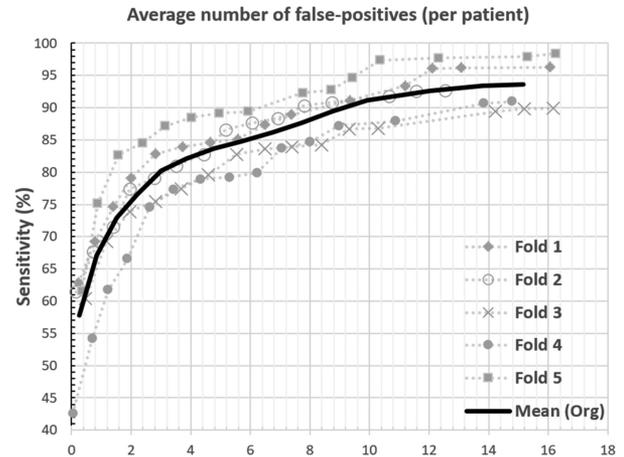


Fig. 9. Average number of false-positives per patient (i.e., wrongly detected BM lesions for each patient) in relation to the sensitivity is illustrated for each CV fold. The mean curve (shown with bold curve) represents the average of the CV folds.

the loss function. For each fold, *CropNet* is trained for 20000 batch iterations, where each batch included 130 pairs of positive and negative samples. The optimal version of the network was determined using the minima of moving validation loss average, computed over 30 batch iterations. On average, the training process took 11312 ($\sigma = 183$) batch iterations to converge. The implementation was performed using Python programming language (v3.6.8) where the neural network was created and trained via Keras library (v2.1.6-tf) with TensorFlow (v1.12.0) backend. The network's training time for each fold was ~ 3.5 hours using an NVIDIA 1080ti graphics card with 11 GB RAM.

The average number of false-positives (i.e., false lesion-detections) per patient (AFP) were computed in connection to the sensitivity of the framework for each CV fold, where the sensitivity of the framework was adjusted via setting a threshold at *CropNet*'s response. AFP was computed as 9.12 per patient with a standard deviation of 3.49 at 90 percent sensitivity. At lower sensitivity percentages, AFP was computed as 8.48 at 89%, 7.92 at 88%, 7.29 at 87%, 6.64 at 86%, and 5.85 at 85% (see Fig. 9). Fig. 11 illustrates sample output screens for the deployed BM detection framework.

To illustrate the impact of the proposed augmentation procedures, the CV study (with the same folds) was performed on (1) the proposed framework where both random elastic deformation and random gamma correction augmentations are excluded – nED-nG, (2) only the random elastic deformation augmentations are excluded – nED, and (3) only the random gamma correction augmentations are excluded – nG (see Fig. 10). For these configurations, AFP values for 80, 82, 85 and 90 sensitivity percentages are reported in Table III.

V. DISCUSSION AND CONCLUSION

The ablation study, performed to visualize the contributions of random elastic deformations and gamma corrections during the augmentation procedure (see Table III), suggests that while both augmentation stages are valuable, positive contribution of random gamma corrections is relatively more prominent; the

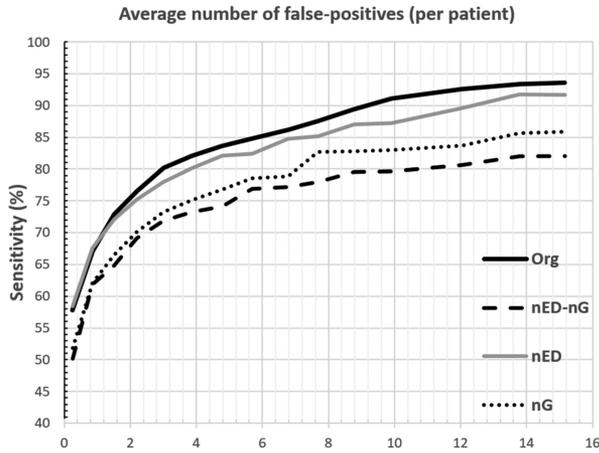


Fig. 10. Average number of false-positives per patient in relation to the sensitivity is illustrated for the proposed solution with all augmentation stages (**Org**), and **nED-nG** (no random elastic deformations or gamma corrections), **nG** (no random gamma corrections) and **nED** (no random elastic deformations) configurations.

TABLE III
AFP VS SENSITIVITY

Sensitivity %	nED-nG	nG	nED	Org
80	10.56	7.05	3.81	2.96
82	14.33	7.52	4.71	3.82
85	-	13.2	7.29	5.85
90	-	-	12.33	9.12

Average number of false positives at specific sensitivity percentages are reported for the proposed solution with all augmentation stages (**Org**), and **nED-nG**, **nG** and **nED** configurations.

framework manages to achieve 90 percent sensitivity with the exclusion of elastic deformations where the AFP value raises to 12.33 (from 9.12 of the original setup). On the other hand, the exclusion of the random gamma corrections sets a limit for the framework's sensitivity at ~ 85 percent (see Fig. 10). Elimination of both augmentations (**nED-nG**) leads to a configuration with ~ 82 percent peak sensitivity while producing AFP of 14.33.

Table IV provides an overview of the databases, acquisition types, neural network architectures, validation strategies and detection accuracies of some of the prominent CNN based BM-detection/segmentation approaches for 3D MRI, published over the recent years. From these, [14] and [18] requires multiple MRI sequences during the BM-segmentation/detection process, whereas [16] benefited from multiple sequences for the training and validation. Our framework trained and validated for a single type of MRI sequence, T1c, and requires only this type of study during its decision-making process.

The dimensional properties of the BM included in a detection study are critical for determining the clinical applicability of a proposed solution. This is due to the fact that smaller lesions are harder to identify even by highly trained neuroradiologists. Consequently, they may greatly benefit from a system trained and validated specifically for that type of data. As illustrated in Table IV, our study employed a BM database that included relatively smaller BM lesions compared with the referenced

studies; the smallest BM average volume in comparable studies is 672 mm^3 [16], whereas the BM average volume in this study is only 159.58 mm^3 .

BM-detection and segmentation databases used in our study and in other comparable studies (as shown in Table IV) are limited with respect to number of cases; they all consist of some hundreds of patients. Estimating the accuracies of such machine learning approaches, trained with a limited amount of datasets, can gain significantly from the usage of CV, as the method minimizes the error of algorithm's predictive performance evaluation [44]. Therefore, we found it valuable to emphasize the validation schemes of comparable studies in Table IV.

The study introduced multiple technical novelties: (1) Sensitivity constrained LoG BM-candidate selection, (2) random 3D Simard elastic deformation augmentations (Simard deformation field used for medical-image augmentation for the first time to our knowledge), (3) volumetric random gamma correction augmentations for MRI, and (4) a parametric CNN for processing cubic volumes of interests. More importantly, all of these components are put into a sound framework that can be utilized for various detection applications in medical imaging.

The limitations of the proposed solution are (1) its scope is currently limited to the detection; medical interventions requiring exact borders of detected tissues (such as stereotactic radiotherapy) may not benefit from the method, and (2) lesions with sizes larger than 15 mm would not be detected with a system trained with the parameters given in this paper. However, BM-candidate selection and *CropNet* parameters can be reset for the dimensional properties of the target tumors.

The performances of machine-learning algorithms, including the CNNs, heavily depend on their hyperparameter settings [45]. Accordingly, some of the BM-segmentation studies, such as [13] and [14], provided a set of analyses on parameter tuning. The introduced framework's performance also relies on proper setup of multiple parameters, including (1) edge length and the block count of *CropNet*, (2) random gamma correction range, and (3) elastic deformation parameters, which were found empirically and individually. Therefore, multivariate optimization of these in a future study might further improve the accuracy of the framework.

The study utilized *CropNet*-b2-16 mm, containing 2 processing blocks per level. Since the number of convolutional layers for the given architecture is significantly low ($\ll 100$), the introduced system is not prone to vanishing/exploding gradients problem as described in [39]. Thus, skip connections in forms of bypassing (e.g. Highway Networks [46], ResNets [47], etc.) or direct paths (e.g. DenseNet [48]) are not part of the given architecture. Therefore, (1) the impact of using high block counts, (2) the architectural enhancements (in forms of skip connections) required to sustain/improve the accuracy level with these deeper architectures, and (3) the validation of this new improvement in connection to the BM detection are topics for a future study.

Transfer learning, enabling the utilization of CNNs pre-trained with relatively large scale image databases (e.g. ImageNet [49]), has been shown to be effective in a variety of imaging applications [50]. However, the CNNs used for transfer

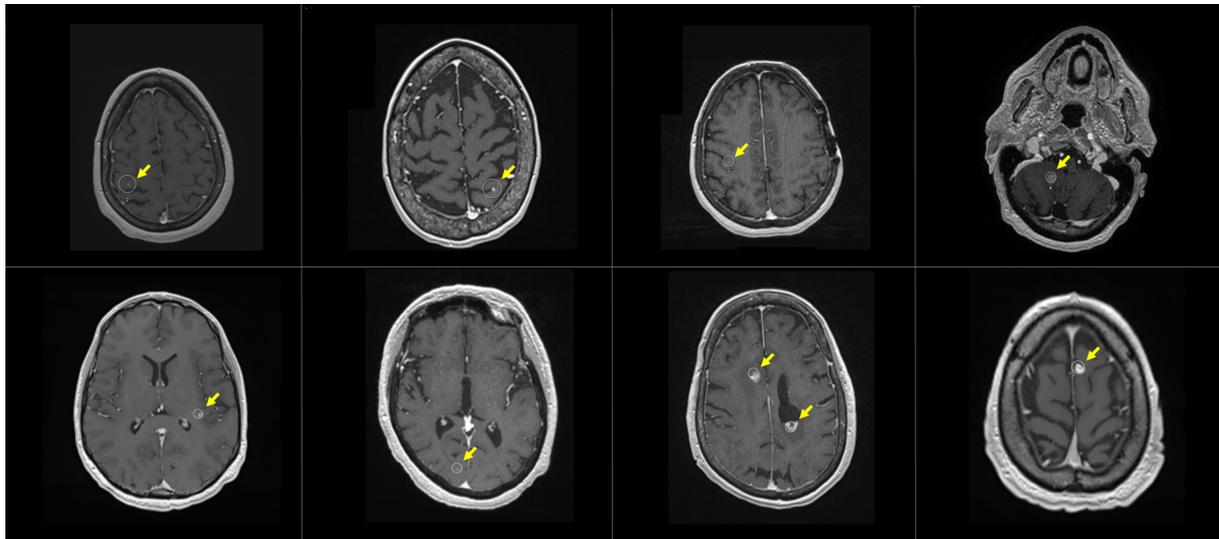


Fig. 11. The framework output; white circles centered by the BM detections are rendered (yellow arrows are added to the figure for the readers' convenience).

TABLE IV
OVERVIEW OF BM DETECTION/SEGMENTATION STUDIES THAT USE CONVOLUTIONAL NEURAL NETWORKS

Study	Patient #	Acquisition	BM diameter (mm)	BM volume (mm)	DNN Type	Validation Type	Sensitivity (%)	AFP
Losch et al. [13]	490	T1c MRI	NA	NA	Multi-scale ConvNet	Fixed train/test ^c	82.8	7.7
Liu et al. [16]	490	Multi seq. ^a	NA	Mean: 672	En-DeepMedic	5-fold CV	NA	NA
Charron et al. [14]	182	Multi seq. ^b	Mean: 8.1 Median: 7	Mean: 2400 Median: 500	DeepMedic	Fixed train/test ^f	93	7.8
Grøvik et al. [18]	156	Multi seq. ^c	NA	NA	GoogLeNet ^d	Fixed train/test ^g	83	8.3
This study	158^h	T1c MRI	Mean: 5.4 Median: 4.6	Mean: 159.6 Median: 50.4	CropNet	5-fold CV	90	9.12

^a235 T1c MRI datasets, and 265 datasets from BRATS DB; including both T1c and T2-weighted Fluid-Attenuated Inversion Recovery (FLAIR) sequences.

^bT1-weighted 3D MRI with Gd injection, T2-weighted 2D fluid attenuated inversion recovery MRI and T1-weighted 2D MRI sequences.

^cPre- and post-gadolinium T1-weighted 3D fast spin echo (CUBE), post-gadolinium T1-weighted 3D axial IR-prepped FSPGR (BRAVO), and 3D CUBE fluid attenuated inversion recovery (FLAIR) sequences.

^d2.5 dimensional fully connected convolutional net based on GoogLeNet.

^e440 training and 50 test cases.

^f164 training and 18 test cases.

^g100 training, 5 development and 51 test cases.

^h217 datasets are collected from 158 patients, CV folds are created patient-wise to ensure a patient can only exist either in training or testing group.

learning tasks are commonly pre-trained with 2D images. Accordingly, in 3D medical imaging, transfer learning is commonly performed via recurrent neural networks (RNNs) [51], which briefly process a given 3D image via slice-by-slice fashion. The applicability of RNNs in the described framework will also be investigated in the future.

As given in Table I, the study was performed on datasets with (1) pixel sizes ranging from 0.43 to 1.0 mm and (2) slice thicknesses ranging from 0.8 to 1.0 mm, where the data was resampled to (1 mm × 1 mm × 1 mm) voxels at preprocessing stage. The results were not compiled for delineating the impact of original pixel size and slice thickness on overall system performance; the validation of the proposed system concerning those

and additional scanner parameters (e.g., imaging frequency, etc.) may also be performed in a future study.

The introduced framework can be extended for segmentation of the metastatic mass lesions. The network's contracting layers can be appended with a symmetric set of expanding layers as in [27] or [25], and its loss function can be changed to Dice similarity coefficient, or another image segmentation metric [52], to perform segmentation. Alternatively, previously defined BM-segmentation algorithms can be modified to use the proposed detection framework in their preprocessing stages.

The proposed data augmentation pipeline uses random gamma transformations and elastic deformations to capture the BM intensity and shape variabilities. The strategy mimics the

kernel density estimation with Parzen windows [53], as the probability densities of the BM with respect to intensity and shape are generated from a small set of actual BM (932 BM) and their ranged uniform variations to deploy a uniform kernel density. For density estimation problems, it is also common to use Gaussian kernel densities [53], which would translate to (1) using gamma corrections randomly picked from a normal distribution centered at 1 (i.e., $\gamma = 1$ gives the original image), and (2) elastic deformations randomly picked from a bivariate distribution centered at (0, 0) (i.e., $\sigma = 0$ and $\alpha = 0$ implies null Simard deformation field). The impact of kernel density function to the final accuracy is a topic for a future study.

This study introduced a novel BM detection framework that focused on small lesions. It is validated for its high sensitivity, and it produced relatively few false BM detections per patient. The results suggest that its detection performance is comparable with state-of-art approaches that are validated against significantly larger lesions. In addition to technical novelties introduced, the study focuses on an increasingly important field-of-research that is the detection of small BM in order to minimize challenges they pose to both radiologists and radiation oncologists in the identification and treatment of these lesions to ultimately improve patient care.

REFERENCES

- [1] L. Nayak, E. Q. Lee, and P. Y. Wen, "Epidemiology of brain metastases," *Curr. Oncol. Rep.*, vol. 14, no. 1, pp. 48–54, 2012.
- [2] B. C. Yoon, A. F. Saad, P. Rezaei, M. Wintermark, G. Zaharchuk, and M. Iv, "Evaluation of thick-slab overlapping MIP images of contrast-enhanced 3D T1-weighted CUBE for detection of intracranial metastases: A pilot study for comparison of lesion detection, interpretation time, and sensitivity with nonoverlapping CUBE MIP, CUBE, a," *Amer. J. Neuroradiol.*, vol. 39, no. 9, pp. 1635–1642, 2018.
- [3] Ú. Pérez-Ramírez, E. Arana, and D. Moratal, "Computer-aided detection of brain metastases using a three-dimensional template-based matching algorithm," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2014, pp. 2384–2387.
- [4] R. D. Ambrosini, P. Wang, and W. G. O'dell, "Computer-aided detection of metastatic brain tumors using automated three-dimensional template matching," *J. Magn. Reson. Imag.*, vol. 31, no. 1, pp. 85–93, 2010.
- [5] R. Farjam, H. A. Parmar, D. C. Noll, C. I. Tsien, and Y. Cao, "An approach for computer-aided detection of brain metastases in post-Gd T1-W MRI," *Magn. Reson. Imag.*, vol. 30, no. 6, pp. 824–836, 2012.
- [6] S. Yang, Y. Nam, M.-O. Kim, E. Y. Kim, J. Park, and D.-H. Kim, "Computer-aided detection of metastatic brain tumors using magnetic resonance black-blood imaging," *Investigative Radiol.*, vol. 48, no. 2, pp. 113–119, 2013.
- [7] H. Khotanlou, O. Colliot, J. Atif, and I. Bloch, "3D brain tumor segmentation in MRI using fuzzy classification, symmetry analysis and spatially constrained deformable models," *Fuzzy Sets Syst.*, vol. 160, no. 10, pp. 1457–1473, 2009.
- [8] D. Dera, N. Bouaynaya, and H. M. Fathallah-Shaykh, "Automated robust image segmentation: Level set method using nonnegative matrix factorization with application to brain MRI," *Bulletin Math. Biol.*, vol. 78, no. 7, pp. 1450–1476, 2016.
- [9] T. Sugimoto, S. Katsuragawa, T. Hirai, R. Murakami, and Y. Yamashita, "Computerized detection of metastatic brain tumors on contrast-enhanced 3D MR images by using a selective enhancement filter," in *Proc. World Congr. Med. Phys. Biomed. Eng.*, 2009, pp. 2070–2072.
- [10] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
- [11] S. Bakas *et al.*, "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge," 2018, *arXiv1811.02629*.
- [12] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, 2017.
- [13] M. Losch, "Detection and segmentation of brain metastases with deep convolutional networks," M.S. thesis, KTH, Comput. Vision and Active Perception, CVAP, 2015.
- [14] O. Charron, A. Lallemand, D. Jarnet, V. Noblet, J.-B. Clavier, and P. Meyer, "Automatic detection and segmentation of brain metastases on multimodal MR images with a deep convolutional neural network," *Comput. Biol. Med.*, vol. 95, pp. 43–54, 2018.
- [15] K. Kamnitsas *et al.*, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Med. Image Anal.*, vol. 36, pp. 61–78, 2017.
- [16] Y. Liu *et al.*, "A deep convolutional neural network-based automatic delineation strategy for multiple brain metastases stereotactic radiosurgery," *PLoS One*, vol. 12, no. 10, 2017, Art. no. e0185844.
- [17] B. H. Menze *et al.*, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Trans. Med. Imag.*, vol. 34, no. 10, pp. 1993–2024, Oct. 2015.
- [18] E. Grøvik, D. Yi, M. Iv, E. Tong, D. Rubin, and G. Zaharchuk, "Deep learning enables automatic detection and segmentation of brain metastases on multi-sequence MRI," *J. Magnet. Reson. Imag.*, vol. 51, no. 1, pp. 175–182, 2020.
- [19] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit. (CVPR)*, 2015, pp. 1–9.
- [20] E. McTyre, J. Scott, and P. Chinnaiyan, "Whole brain radiotherapy for brain metastasis," *Surgical Neurol. Int.*, vol. 4, no. Suppl 4, 2013, Paper S236.
- [21] T. Lindeberg, "Image matching using generalized scale-space interest points," *J. Math. Imag. Vis.*, vol. 52, no. 1, pp. 3–36, 2015.
- [22] T. Lindeberg, "Scale selection properties of generalized scale-space interest point detectors," *J. Math. Imag. Vis.*, vol. 46, no. 2, pp. 177–210, 2013.
- [23] C.-P. Yu, G. Ruppert, R. Collins, D. Nguyen, A. Falcao, and Y. Liu, "3D blob based brain tumor detection and segmentation in MR images," in *Proc. IEEE 11th Int. Symp. Biomed. Imag.*, 2014, pp. 1192–1197.
- [24] S. Wang and R. M. Summers, "Machine learning and radiology," *Med. Image Anal.*, vol. 16, no. 5, pp. 933–951, 2012.
- [25] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," 2016, pp. 565–571.
- [26] A. Dosovitskiy, P. Fischer, J. T. Springenberg, M. Riedmiller, and T. Brox, "Discriminative unsupervised feature learning with exemplar convolutional neural networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 9, pp. 1734–1747, Sep. 2016.
- [27] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervent.*, 2015, pp. 234–241.
- [28] X. Yang *et al.*, "Co-trained convolutional neural networks for automated detection of prostate cancer in multi-parametric MRI," *Med. Image Anal.*, vol. 42, pp. 212–227, 2017.
- [29] J. Duchon, "Splines minimizing rotation-invariant semi-norms in Sobolev spaces," in *Constructive Theory of Functions of Several Variables*. Berlin, Germany: Springer, 1977, pp. 85–100.
- [30] M. H. Le *et al.*, "Automated diagnosis of prostate cancer in multi-parametric MRI based on multimodal convolutional neural networks," *Phys. Med. Biol.*, vol. 62, no. 16, 2017, Paper 6497.
- [31] P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *Proc. 7th Int. Conf. Document Anal. Recognit.*, 2003, vol. 1, pp. 958–963.
- [32] N. Robitaille, A. Mouiha, B. Crépeault, F. Valdivia, and S. Duchesne, "Tissue-based MRI intensity standardization: Application to multicentric datasets," *J. Biomed. Imag.*, vol. 2012, 2012, Art. no. 347120.
- [33] M. Shah *et al.*, "Evaluating intensity normalization on MRIs of human brain with multiple sclerosis," *Med. Image Anal.*, vol. 15, no. 2, pp. 267–282, 2011.
- [34] J. D. Christensen, "Normalization of brain magnetic resonance images using histogram even-order derivative analysis," *Magn. Reson. Imag.*, vol. 21, no. 7, pp. 817–820, 2003.
- [35] B. M. Ellingson *et al.*, "Comparison between intensity normalization techniques for dynamic susceptibility contrast (DSC)-MRI estimates of cerebral blood volume (CBV) in human gliomas," *J. Magn. Reson. Imag.*, vol. 35, no. 6, pp. 1472–1477, 2012.
- [36] L. G. Nyúl, J. K. Udupa, and X. Zhang, "New variants of a method of MRI scale standardization," *IEEE Trans. Med. Imag.*, vol. 19, no. 2, pp. 143–150, Feb. 2000.

- [37] S. Roy, A. Carass, and J. L. Prince, "Patch based intensity normalization of brain MR images," in *Proc. IEEE 10th Int. Symp. Biomed. Imag.*, 2013, pp. 342–345.
- [38] A. Galdran *et al.*, "Data-driven color augmentation techniques for deep skin image analysis," 2017, *arXiv:1703.03702*.
- [39] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.
- [40] M. Demirer *et al.*, "A user interface for optimizing radiologist engagement in image-data curation for artificial intelligence," *Radiol. Artif. Intell.*, vol. 1, no. 6, doi: <https://doi.org/10.1148/ryai.2019180095>.
- [41] J. Pluim, J. Maintz, and M. Viergever, "Mutual-information-based registration of medical images: A survey," *Med. Imag., IEEE Trans.*, vol. 22, no. 8, pp. 986–1004, Aug. 2003.
- [42] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.
- [43] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, 2014, *arXiv:1412.6980*.
- [44] B. J. Erickson, P. Korfiatis, Z. Akkus, and T. L. Kline, "Machine learning for medical imaging," *Radiographics*, vol. 37, no. 2, pp. 505–515, 2017.
- [45] J. N. van Rijn and F. Hutter, "Hyperparameter importance across datasets," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2018, pp. 2367–2376.
- [46] R. K. Srivastava, K. Greff, and J. Schmidhuber, "Training very deep networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 2377–2385.
- [47] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2016, pp. 770–778.
- [48] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, pp. 4700–4708.
- [49] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2009, pp. 248–255.
- [50] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *J. Big Data*, vol. 3, no. 1, pp. 1–40, 2016.
- [51] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep learning techniques for medical image segmentation: Achievements and challenges," *J. Digit. Imag.*, vol. 32, pp. 1–15, 2019.
- [52] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool," *BMC Med. Imag.*, vol. 15, no. 1, p. 29, 2015.
- [53] T. Hastie, R. Tibshirani, and J. Friedman, "Kernel Density Estimation and Classification," in *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Berlin, Germany: Springer, 2009, pp. 208–212.