Retraction

Retracted: Using Convolutional Neural Networks for Segmentation of Multiple Sclerosis Lesions in 3D Magnetic Resonance Imaging

Advances in Materials Science and Engineering

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] A. Abdullah Hamad, M. Musa Jaber, M. Altaf Ahmed, G. Muttashar Abdulsahib, O. Ibrahim Khalaf, and Z. Meraf, "Using Convolutional Neural Networks for Segmentation of Multiple Sclerosis Lesions in 3D Magnetic Resonance Imaging," *Advances in Materials Science and Engineering*, vol. 2022, Article ID 4905115, 10 pages, 2022.

Hindawi

Research Article

Using Convolutional Neural Networks for Segmentation of Multiple Sclerosis Lesions in 3D Magnetic Resonance Imaging

Abdulsattar Abdullah Hamad [1,2] **Mustafa Musa Jaber,**[3,4] **Mohammed Altaf Ahmed** [5] **Ghaida Muttashar Abdulsahib,**[6] **Osamah Ibrahim Khalaf,**[7] **and Zelalem Meraf** [8]

[1]*Department of Medical Laboratory Techniques, Dijlah University College, Baghdad 10021, Iraq*
[2]*Department of Medical Laboratory Techniques, Al-Turath University College, Baghdad 10021, Iraq*
[3]*Department of Medical Instruments Engineering Techniques, Dijlah University College, Baghdad 10021, Iraq*
[4]*Department of Medical Instruments Engineering Techniques, Al-Farahidi University, Baghdad 10021, Iraq*
[5]*Department of Computer Engineering, College of Computer Engineering & Sciences, Prince Sattam Bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia*
[6]*Department of Computer Engineering, University of Technology, Baghdad, Iraq*
[7]*Al-Nahrain Nanorenewable Energy Research Center, Al-Nahrain University, Baghdad, Iraq*
[8]*Department of Statistics, Injibara University, Injibara, Ethiopia*

Correspondence should be addressed to Zelalem Meraf; zelalemmeraf@inu.edu.et

Magnetic Resonance Imaging to detect its lesions is used to diagnose multiple sclerosis. Experts usually perform this detection process manually, but there is interest in automating it to speed up the diagnosis and monitoring of this disease. A variety of automatic image segmentation methods have been proposed to quickly detect these lesions. A Gaussian Mixture Model is first constructed to identify outliers in each image. Then, using a set of rules based on expert knowledge of multiple sclerosis lesions, those outliers of the model that do not match the lesions' characteristics are discarded. Furthermore, segmented lesions usually correspond to gray matter-rich brain regions. In some cases, false positives can be detected, but the rules used cannot eliminate all errors without jeopardizing the segmentation's quality. The second method involves training a convolutional neural network (CNN) that can segment lesions based on a set of training images. This technique can learn a set of filters that, when applied to small sections of an image called "patches," produce a set of characteristics that can be used to classify each voxel of the image as a lesion or healthy tissue. On the other hand, the results show that the networks are capable of producing results in the worked database comparable to those produced by the algorithms in the literature.

1. Introduction

Magnetic resonance imaging (MRI) is a medical imaging technique that allows obtaining images of different parts of the body, such as the brain, heart, or spine, helping diagnose various pathologies that may be affecting a patient [1]. This technique has many advantages over other brain scanning techniques, including the use of nonionizing radiation and the generation of images with a high degree of contrast between the different types of tissues present in the brain (cerebrospinal fluid, white and gray matter), which makes it an excellent support for diagnosing diseases that affect the central nervous system [2], such as epilepsy and dementia.

Among these diseases is multiple sclerosis, a chronic disease that causes damage to the central nervous system, characterized by the appearance of regions of inflammation and demyelination of neurons, which result in a progressive loss of the patient's motor, sensory, and cognitive abilities

affected. Through MRI, it is possible to detect the regions affected by this disease and make a diagnosis based on the number and size of the lesions detected [3].

This task is generally performed by an expert medical professional or semiautomatically, using computational image segmentation techniques [4], which require parameter adjustment by the expert. This approach tends to present difficulties and disadvantages, such as high variability in the results issued by different experts. Due to this, the medical community's interest arises in developing automatic segmentation techniques to reduce the time required to identify damaged regions in the brain. There are many challenges in automatic segmentation problems, and their use in practice is quite limited due to the difficulty of correctly validating the predictions made by these tools. In this memory, the effectiveness of different segmentation methods for this type of image will be studied, proposing the use of convolutional neural networks for the segmentation of lesions and comparing them with traditional segmentation methods, analyzing the advantages and disadvantages that each of these methods they bring into practice [5].

This work aims to present different techniques of the literature in the area of image segmentation. Then, the implementation of these and their comparison in terms of performance and effectiveness in the segmentation problem of magnetic resonance images will be carried out. To validate results, the set of Boston Children's Hospital (BCH) training images of the MS Lesion Segmentation Challenge will be used [6], which consists of 10 images in T1, T2, and FLAIR modalities of different patients, accompanied by the segmentation of an expert. Because the segmentation of the test set is not publicly accessible, for supervised learning algorithms, in this case, convolutional neural networks, the set of images will be divided into a training subset, from which the model will be trained of convolutional networks, and a test set, with which the ability of the model to segment lesions in a new image correctly will be tested. This process will be performed five times per technique and an average of the performance measurements will be obtained. The current study aims to experimentally evaluate the performance of convolutional neural networks applied to the problem of automatic segmentation of multiple sclerosis (MS) lesions in multimodal magnetic resonance imaging, comparing it with other literary techniques.

2. Methodology

The proposed solution to the problem of segmentation of multiple sclerosis lesions in magnetic resonance images consists of implementing and evaluating two different methods applied in the literature to the problem. The dataset with which the proposal will be validated corresponds to the database of training images used in the multiple sclerosis lesion segmentation challenge organized by the Medical Image Computing and Computer-Assisted Intervention Society [7]. This database is available to the public and comprises magnetic resonance images of twenty different patients in 3 different modalities: T_1-Weighted, T_2-

Weighted, and T_2-Fluid Attenuated Inversion Recovery (FLAIR), with their respective masks of segmented lesions. Ten of these images in their 3 modalities, with their respective segmentations, were provided by the University of North Carolina (UNC). The other ten were provided by Children's Hospital Boston (BCH).

The modalities of each image of the data set are coregistered, that is, the voxels of each image correspond to the same spatial coordinate of the brain. However, they lack a greater preprocessing, for which a preprocessing stage is necessary before the analysis application of segmentation algorithms to obtain good results.

Once the images have been preprocessed, two literature segmentation algorithms will be implemented. Once state of the art was evaluated, it was decided to implement a technique based on the Gaussian Mixture Models [8] based on the search for outliers in the model and the use of rules to eliminate false positives from the segmentation. On the other hand, the implementation of a convolutional neural network was chosen to segment lesions [9]. This method has become very popular in the literature in recent years.

To compare the effectiveness of both methods, the Sensitivity (True Positive Rate) and Precision (Positive Predictive Value) metrics will be used, which are given by

$$
\begin{aligned}
\text{TPR} &= \frac{TP}{TP + FN}, \\
\text{PPV} &= \frac{TP}{TP + FP},
\end{aligned}
\tag{1}
$$

where TP is the number of true positives from segmentation (correctly labeled lesions), FP is the number of false positives from segmentation (healthy regions that were labeled as lesions by segmentation), and FN is the number of false negatives from segmentation (lesions that were labeled as healthy tissue). Each of these measurements is quantified at the voxel level. Sensitivity represents the percentage of the correctly segmented lesion, while Precision measures the percentage of hits concerning the segmented region. A high Sensitivity indicates that most of the lesion voxels are correctly segmented. In contrast, a low Sensitivity indicates that lesions are not being detected or only a tiny portion of them are being detected. On the other hand, a high Precision indicates that a tiny percentage of the voxels segmented as lesions correspond to healthy tissue. In contrast, a low Precision indicates that the percentage of healthy tissue segmented as lesions is very high due to a large number of false positives. Finally, the Sorensen Dice similarity (DSC) [10] coefficient will be used, a standard metric used in the literature that allows determining the level of overlap between the obtained segmentation and the segmentation provided by an expert, which is given by the harmonic mean between the two metrics.

$$
\begin{aligned}
\text{DSC} &= \frac{2}{1/\text{TPR} + 1/\text{PPV}}, \\
&= \frac{2TP}{2TP + FP + FN}.
\end{aligned}
\tag{2}
$$

3. Construction and Implementation of the Solution

3.1. Preprocessing. To obtain good results in the segmentation algorithms, it was necessary to perform a preprocessing stage to eliminate image noise and other irregularities present.

3.1.1. Bias Field Correction. MRI images fall victim to a signal that corrupts image quality known as the bias field. This signal is the product of inhomogeneities in the magnetic field used to obtain these images, and its presence results in blurred sectors in the image, adding noise to the intensity values of the image, which negatively affects the quality of the results obtained by algorithms of image processing and segmentation since these disturbances corrupt the distribution of intensities of the different objects and regions that appear in the image. To correct this signal, the N4ITK algorithm was used.

The N4ITK algorithm consists of a modification of the N3 algorithm (nonparametric nonuniformity normalization) that seeks to resolve some algorithm's shortcomings [11]. Given an image, the algorithm models the bias field as a function that multiplies the image without corruption as

$$v(x) = u(x)f(x) + n(x), \tag{3}$$

where $v(x)$ is the original image; $u(x)$ is the image without corruption; $f(x)$ is the bias field; $n(x)$ is the noise (assumed to be independent Gaussian noise).

Using the notation $\hat{u} = \log u$, the previous equation applying logarithm is of the form:

$$\hat{v}(x) = \hat{u}(x) + \hat{f}(x). \tag{4}$$

From this point, the image without corruption is iteratively estimated by

$$\begin{aligned}\hat{u}^n &= \hat{u}^{n-1} - \hat{f}_r^n, \\ &= \hat{u}^{n-1} - S^*\{\hat{u}^{n-1} - E[\hat{u}|\hat{u}^{n-1}]\},\end{aligned} \tag{5}$$

where $S^*\{\bullet\}$ is an approximation by B-Splines of the bias field residual between the nth and $n-1$th iterations. Details of this implementation can be found in [11].

3.1.2. Skull Stripping. To avoid involving the voxels of the images corresponding to the background and the skull, the skull stripping task was performed to obtain a binary mask that only included those voxels corresponding to brain tissues [12]. For this task, BrainSuite software was used. The skull stripping task was performed using the BSE (Brain Surface Extractor) algorithm. Parameter selection was performed manually for each image until an appropriate mask was obtained that did not include skull voxels. The parameters that produced satisfactory results for each of the images in the BCH database were seen in Figure 1 and Table 1.

3.2. Gaussian Mixture Models (GMM). As in Karimian and Jafari [13] paper, the first method chosen for segmentation is modeling the intensities of the three types of brain tissue using 3-component Gaussian Mixture Models. The segmentation will be done by looking for outliers in the model using the truncated maximum likelihood estimator of the GMM, then applying a set of rules to eliminate false positives.

The algorithm proposed by Lee and Song [14] is dependent on several parameters, including the GMM's percentage of outliers and the thresholds used to eliminate them. To choose these parameters, we vary the percentage of outliers in the model h between {0; 0.2; 0.3; 0.4; 0.5}. Then, the Mahalanobis distance, T2 image intensity, and FLAIR image intensity thresholds were varied for each model. The Mahalanobis distance ranged from 0.4 to two. The parameters chosen correspond to the highest DSC combination.

A problem that was observed when adjusting the GMMs for the different images in the database is that the intensity distributions for white matter in the T2 and FLAIR images varied a lot between the different images, for which the use of a threshold single fixed for T2 and FLAIR images did not produce good results in segmentation. To solve this problem, an idea similar to that proposed in Jansen [15] of using thresholds based on parameters of the obtained model was taken. In this way, the regions of hyperintensity in the white matter of the T2 and FLAIR images are selected by the thresholds μWM, $T2 + kT2 \bullet \sigma$WM, T2 and μWM, FLAIR + kFLAIR $\bullet \sigma$WM, FLAIR, where μWM, i is the mean of the white matter intensities in the image in modality i, σWM, i is the standard deviation of the white matter intensities in the image in modality i, and kT2 and kFLAIR are 2 constants associated with the threshold of their respective image. These last 2 parameters replace the fixed threshold defined by Karimian and Jafari and were made to vary in the range of values {0, 1, 2}. Finally, hypointensity in the T1 image, defined by the negative threshold μWM, T1 – kT1 $\bullet$ σWM, T1, was also added to the set of rules. The parameter kT1 was varied in the same range as kT2 and kFLAIR.

Another issue with the method described in the paper is that most predicted lesions were located at the outer edge of the white matter, near the cortical gray matter. The solution proposed by Zhao et al. [16] is to apply a mask obtained from the morphological erosion operation to the white matter mask, allowing the elimination of nearby voxels from the lesion candidates. The internal holes of the white matter mask were filled before applying the erosion mask. The chosen erosion filter was $6 \times 6 \times 6$. A 3×3 erosion filter was applied to each slice of their set of images, whose resolution was half that of the images with which they worked, which is why the filter size was initially set to 6 6 6. These voxels were segmented as a lesion and their final size was set to $9 \times 9 \times 9$, see Table 2.

The initial segmentation to obtain the estimators of the initial parameters of the model was performed on the T1 image using the *K*-means algorithm.
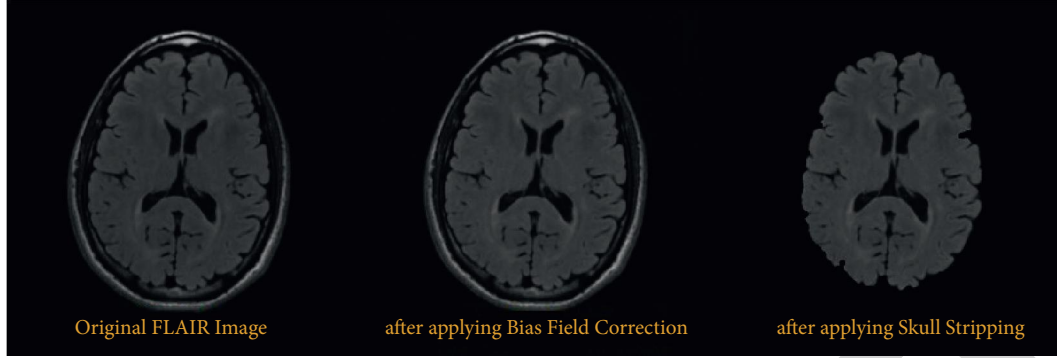
Figure 1: Original FLAIR image, image after applying bias field correction, and image after applying skull stripping.

Table 1: Parameters used in the BCH.

Image	Edge constant	Erosion size
BCH01	0.7	2
BCH02	0.7	2
BCH03	0.7	2
BCH04	0.9	2
BCH05	0.6	3
BCH06	0.7	3
BCH07	0.7	2
BCH08	0.7	3
BCH09	0.7	3
BCH10	0.7	3

Table 2: Parameters of means obtained for the white matter.

Image no.	μWM, T2	μFLAIR, T2
BCH01	77.89	73.7
BCH02	37.76	62.18
BCH03	38.08	80.85
BCH04	46.4	100.96
BCH05	45.26	63.65
BCH06	50.01	61.64
BCH07	53.88	59.78
BCH08	43.88	60.7
BCH09	58.38	57.84
BCH10	58.33	56.88

3.2.1. K-Means Algorithm. The *K*-means algorithm is an unsupervised learning algorithm commonly applied in clustering tasks. Given a set of points $X = \{x_1, x_2, \ldots, x_n\}$ and a parameter k, corresponding to a positive integer, the algorithm is initialized by choosing k random points $c_1, c_2, \ldots, c_k$ within the same domain of X known as centroids. Then, for each point within the set X, the distance between xi and each of the centroids cj is calculated (commonly, the Euclidean distance is used) and the point xi is assigned to the cluster corresponding to the closest centroid. Finally, the centroids are calculated again as

$$C_j = \frac{1}{n_j} \sum_{x_i \in C_j} x_i, \tag{6}$$

where C_j is the cluster containing the previously closest points to C_j and n_j is the number of points in the cluster C_j. These 2 steps are repeated iteratively until the cluster

assignment for each point does not change from one iteration to the next.

This algorithm applies to the image segmentation task. Given the set of pixels of a grayscale image, the points to be grouped correspond to the intensities of the pixels. In contrast, the centroids correspond to different intensity values in the range [0, 255]. As a result, the *K*-means algorithm performs a grouping of pixels with similar intensity values. The *K*-means algorithm has been used to perform the segmentation of different brain tissues in magnetic resonance images [17].

The K-means algorithm was used to perform a preliminary segmentation of the brain in the 3 different types of tissue based on the intensity of the pixels in the T1 image, obtaining, from lower to higher intensity, labels for the pixels corresponding to cerebrospinal fluid, spinal fluid, and gray and white matter.

3.3. Convolutional Neural Networks (CNNs). Second, the first Birenbaum architecture [18] is used to train a convolutional network using a subsampling of FLAIR image patches. Instead of a longitudinal MRI, the paper's SISTP (Single Image, Single TimePoint) architecture was used. In addition to the proposed architecture, an architecture without the Max pooling layer following the convolution layers was tested to compare performance. Both architectures tested 2 input types: 1 axial image and 1 image for each view type (axial, sagittal, and coronal). Each slice was obtained with a resolution of 0.5 mm and a distance of 1 mm between each slice. The intermediate values were interpolated during the registration process to obtain the final image with a resolution of 0.5 mm. It is then passed to a dense network of 16 neurons followed by the Softmax layer, which concatenates the 48 feature vectors obtained after the first dense layer of each image.

The implementation of the network was carried out using Keras [19], a library written in Python for the management and construction of neural networks in a modular way, using the Tensorflow library [20] as a backend.

In Figure 2, candidate voxels were chosen based on preliminary white matter segmentation from the T1 image using the *K*-means algorithm. This white matter mask replaced the segmentation obtained from registering an Atlas to a T1 image. WM, FLAIR + $k_{FLAIR} \bullet$ WM, $_{FLAIR}$ was
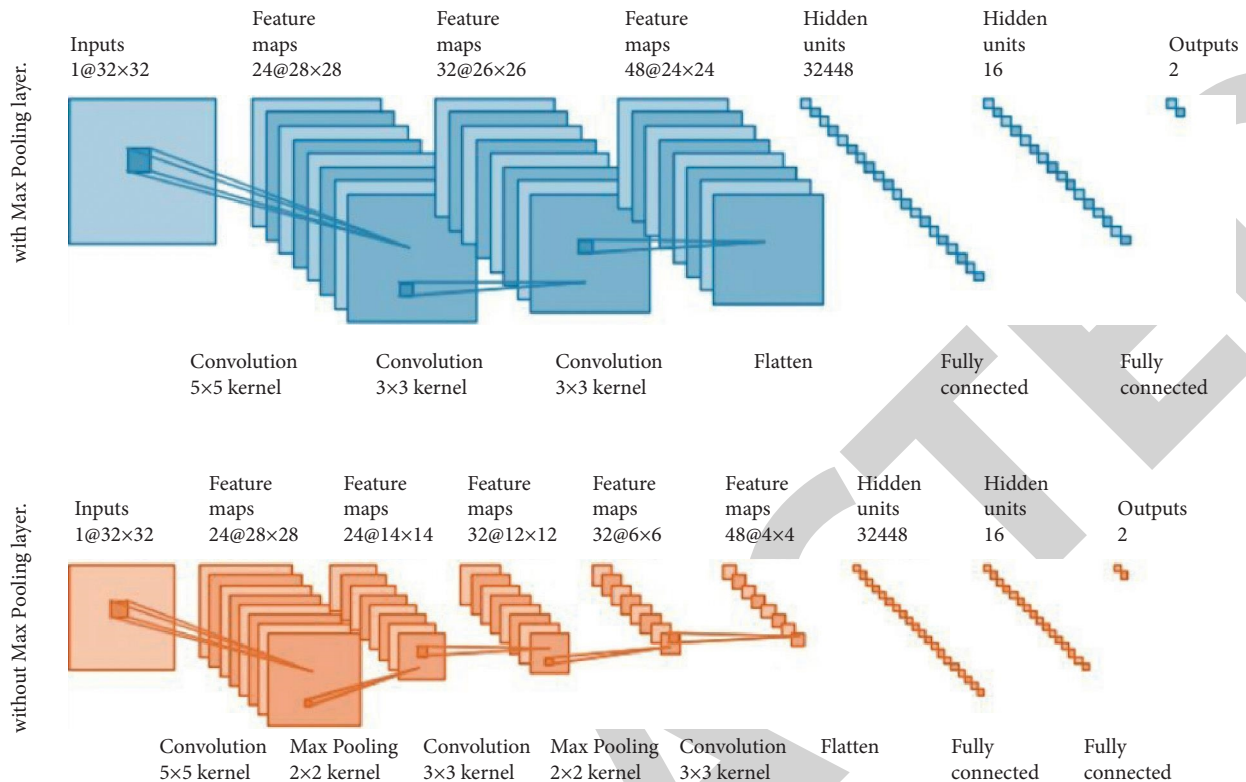
FIGURE 2: Architecture for 1 implemented input of the CNN with and without Max pooling layer.

used to define the threshold parameter for the FLAIR image [21]. The parameter was set to match the proportion of candidate voxels in Birenbaum et al. The final number of candidates for the network was 1,242,109 voxels (kFLAIR = 2), which is close to the percentage reported in Birenbaum's paper, who claims that candidate sampling reduced network computations by 97.5 percent [20]. This subsampling leaves out a percentage of the image voxels corresponding to lesions. In addition to patch sampling, a loss function that penalized false negatives for lesions addressed the class imbalance issue. The weight parameter of this loss function was set at 20% for all trained networks based on the percentage of injuries observed in the training set. We will use leave-one-out cross-validation to evaluate the network's results. The network's model is then applied to the untrained image, yielding model evaluation metrics for that image. This is done for every image in the training set. Finally, the obtained model's evaluation metrics are averaged.

4. Results and Discussion

The results of the selection of parameters for the GMM are presented in the following graph (3), detailing the average DSC obtained for each parameter throughout all the images in the database. The best combination of parameters found, with respect to the DSC obtained, was $h = 0.2$; $d_{min} = 0.4$; $k_{T2} = 0$, and $k_{FLAIR} = 1$.

A larger area of the brain is considered a candidate for injury, so the number of true positives increases. There is a segmentation of more false positives, only a fraction of

which are eliminated by the false positive reduction rules. The h parameter also has the issue of losing information in the GMM due to training with fewer voxels, resulting in poor initial segmentation results. The elimination rule for voxels near the gray matter border reduced the number of false positives because many of the outliers detected in the model were in this region and were not filtered by the lesion selection rules. Figure 3 compares the Positive Predictive Value before and after applying this rule; see Figure 4.

Because the original papers from which the implemented techniques came used different databases to validate his experiments, the results obtained by other techniques were investigated in the public database of the MS Lesion Segmentation Challenge. Overall, the results show a high PPV for most of the techniques used, with relatively precise segmentation of multiple sclerosis lesions. The tradeoff is that the TPR obtained was generally low compared to the state-of-the-art techniques. As for the GMM segmentation, this is due to applying a set of rules that filter out false positives and in some cases remove voxels that corresponded to injuries. The low number of true positives in convolutional networks is due to subsampling used to reduce network training times. The main advantage of GMM is the speed of segmentation. The average time required to segment each image was 538.61 15.47 seconds, including reading the file and applying the EM algorithm to the model. Across all test cases, the EM algorithm took 27.96 3.04 seconds per iteration. Convolutional networks performed better on average, with similarity coefficients of 0.33 and 0.29 for 1-view networks and 0.34 and 0.26 for multiview networks, comparable to
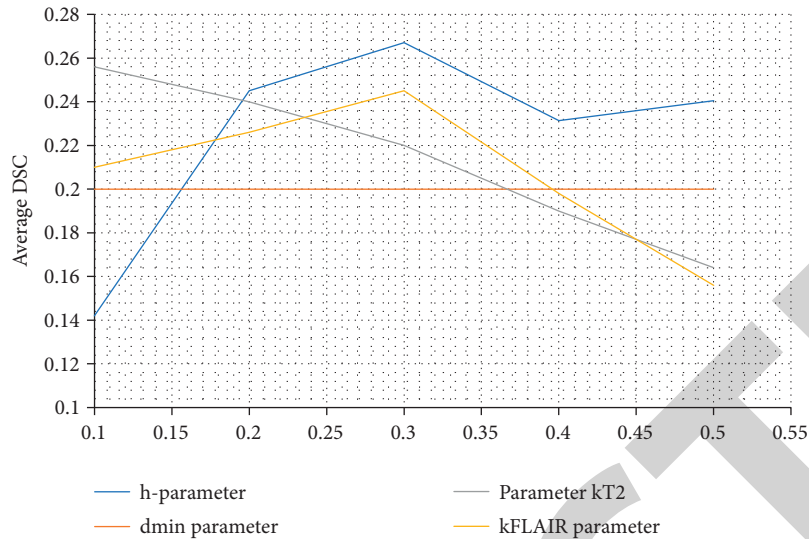
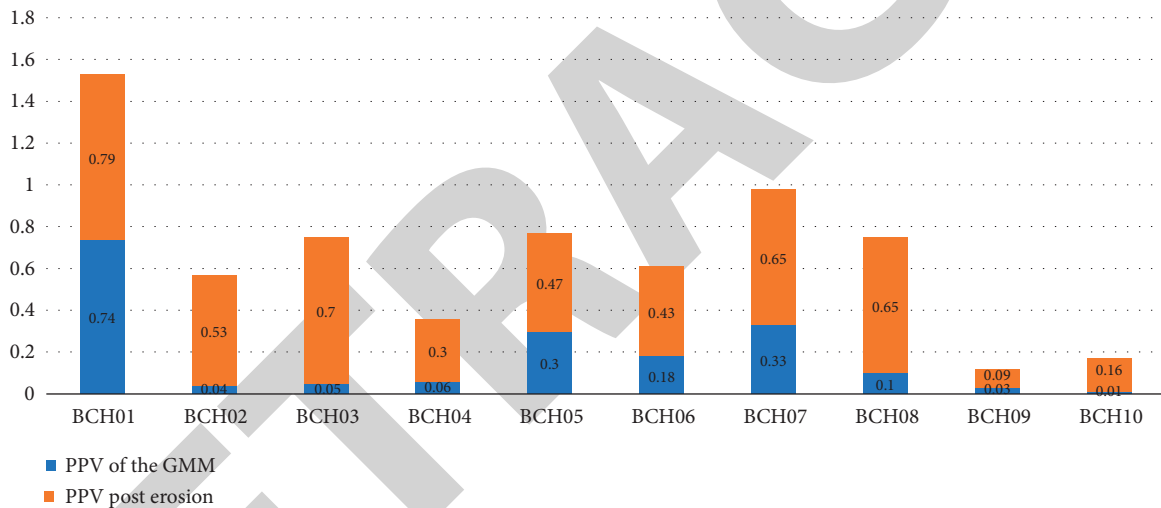Figure 3: The results of the selection of parameters for the GMM.

Figure 4: Comparison of PPV before and after application of erosion to the white matter mask.

most results in the state of the art. The network's average performance was better without the Max pooling layer in terms of results quality. The network underestimates the size of lesions (which explains the low TPR) and simply fails to identify some injuries when the Max pooling layer is added. The number of lesions detected, that is, those for which at least one voxel was classified as a lesion, was as follows.

In Figure 5, the network with the Max pooling layer was able to detect only a subset of the lesions that the network without it could detect. The only exception to this was image BCH08, in which the network with Max pooling for the axial view managed to find 2 lesions that were not detected by the other network; however, these successes could be classified as noise simply by looking at the results of the segmentation because they appear only in 1 or 2 slices of the image and cover a very small region.

In addition, in these images, it can be seen that the segmented regions corresponding to a lesion are not necessarily contiguous, and multiple segmented lesions may appear that correspond to a single lesion see Table 3.

It can be seen that the number of regions that are segmented by the CNN is quite large, which is observed in the results as a large number of small-sized regions that appear in the segmentation around larger regions (usually well-defined lesions) classified), regions around the lateral ventricles of the brain and at the outer edges of the white matter, on the border with the gray matter. An alternative, studied later, to mitigate this discrepancy problem consists of performing a postprocessing stage in which morphological operations are applied to the segmentation mask, improving its quality.

Another problem observed with the segmentation of networks is the dependence on the initial selection of candidates. Although much of the white matter approximation using K-means yielded good results, there are some cases in which the hypointense areas in the white matter were classified as another type of tissue.
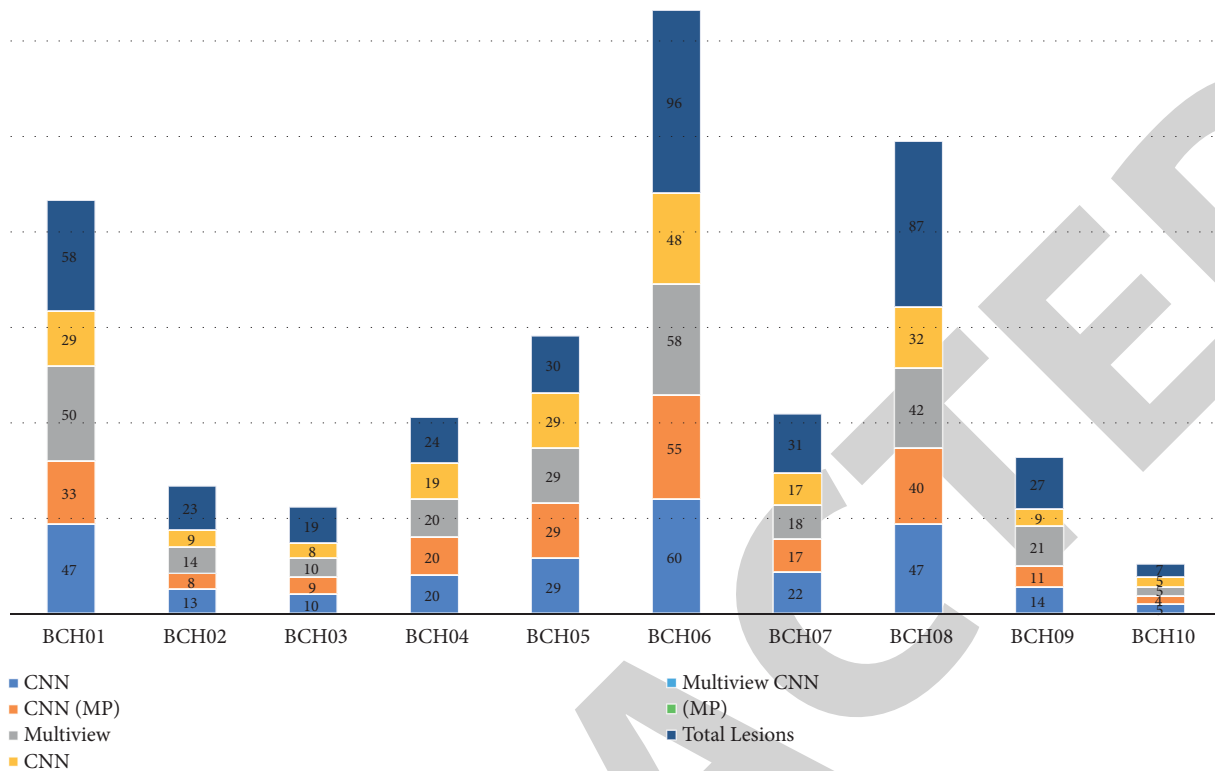
Figure 5: Number of segmented lesions.

Table 3: Number of segmented regions.

Image	CNN		CNN (max pooling)		Multiview CNN		Multiview CNN (MP)	
	No. of regions	Size >6	No. of regions	Size >6	No. of regions	Size >6	No. of regions	Size >6
BCH01	716	173	401	111	536	128	519	115
BCH02	197	57	148	40	414	103	491	122
BCH03	482	120	434	111	465	112	197	54
BCH04	1708	280	2084	300	1374	217	826	162
BCH05	1218	237	693	150	907	187	399	92
BCH06	670	165	512	149	599	156	352	115
BCH07	511	108	405	97	370	86	397	100
BCH08	545	129	465	117	420	104	384	86
BCH09	358	75	386	95	626	124	160	40
BCH10	159	38	133	32	309	80	60	20

In this case, the GMM segmentation is more accurate because it considers all tissue types when determining injury candidates. Although the GMM, like *K*-means, failed to segment white matter, all voxels were considered in the selection of outliers allowing better detection of lesions, not just those predicted to be white matter. One of the benefits of using the Max pooling layer is the reduced training time. The Max pooling layer reduced the feature map size, reducing the network training time by approximately 3. The results of multiview networks were not recorded due to the lengthy training time on the CPU (in hours, minutes, and seconds of execution). The GPU (Graphic Processing Units) reduces the training time. The time required to train the networks, even on the GPU, exceeds the average time required to obtain results from the GMM, which was approximately 9 minutes per image. This time includes file reading,

algorithm execution, and segmentation generation. Unlike convolutional networks, which can take days to train on the CPU, the GMM algorithm can be executed on the CPU without compromising execution time; see Figure 6.

4.1. Postprocessing. To improve the quality of the segmentation, morphological operations were applied to the obtained segmentations. The operation performed consists of a closure, which consists of applying a dilation filter followed by applying an erosion filter. The size of this filter was set at 5×5 to connect those regions that are close to each other but not enough to join regions that are not related to each other.

The results obtained after applying this filter, in addition to the percentage change observed for each metric, are found in Table 4.
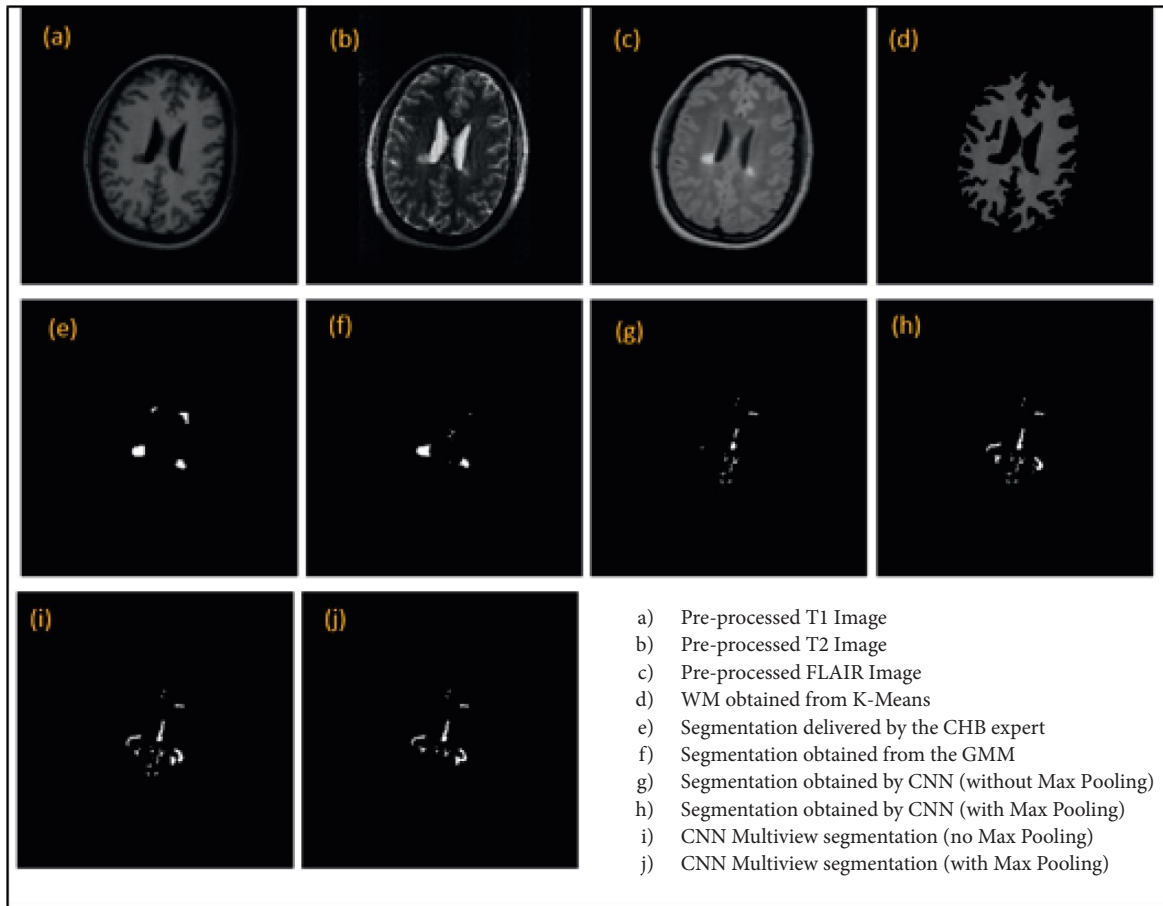
Figure 6: Example of an image in which the automatic segmentation of WM fails. (a) Preprocessed T1 image. (b) Preprocessed T2 image. (c) Preprocessed FLAIR image. (d) WM obtained from *K*-means. (e) Segmentation delivered by the CHB expert. (f) Segmentation obtained from the GMM. (g) Segmentation obtained by CNN (without max pooling). (h) Segmentation obtained by CNN (with Max pooling). (i) CNN multiview segmentation (no Max pooling). (j) CNN multiview segmentation (with Max pooling).

Table 4: Postprocessing results.

Image no.	GMM			CNN			CNN			(Max pooling)			Multiview CNN			Multiview CNN (MP)		
	TPR	PPV	DSC	TPR	PPV	DSC	TPR	PPV	DSC	TPR	PPV	DSC	TPR	PPV	DSC	TPR	PPV	DSC
BCH01	0.46	0.8	0.59	0.58	0.57	0.57	0.33	0.6	0.42	0.55	0.59	0.57	0.32	0.64	0.32
BCH02	0.14	0.53	0.23	0.43	0.61	0.5	0.19	0.79	0.3	0.46	0.56	0.5	0.26	0.55	0.35
BCH03	0.2	0.69	0.3	0.39	0.47	0.43	0.35	0.49	0.41	0.4	0.46	0.43	0.25	0.65	0.36
BCH04	0.44	0.26	0.33	0.16	0.11	0.13	0.38	0.16	0.23	0.32	0.18	0.23	0.34	0.29	0.32
BCH05	0.2	0.63	0.3	0.25	0.21	0.23	0.26	0.26	0.26	0.25	0.23	0.24	0.23	0.32	0.26
BCH06	0.12	0.41	0.19	0.3	0.58	0.39	0.27	0.62	0.37	0.29	0.6	0.39	0.24	0.66	0.35
BCH07	0.34	0.62	0.44	0.37	0.69	0.48	0.32	0.73	0.44	0.36	0.73	0.48	0.28	0.77	0.41
BCH08	0.27	0.59	0.37	0.42	0.77	0.55	0.31	0.77	0.44	0.41	0.78	0.55	0.26	0.77	0.39
BCH09	0.04	0.07	0.05	0.24	0.53	0.33	0.3	0.53	0.3	0.3	0.42	0.35	0.11	0.61	0.19
BCH10	0.02	0.12	0.04	0.09	0.72	0.16	0.07	0.61	0.12	0.13	0.61	0.23	0.04	0.79	0.07
Average	0.22	0.47	0.28	0.32	0.53	0.38	0.28	0.56	0.33	0.35	0.52	0.40	0.23	0.61	0.30

The quality of segmentation improved significantly for all methods. That is because using morphological closure helped to alleviate one of the main issues with segmentations, which was an underestimation of the lesions' size, which was often mistaken for multiple isolated regions. The dilation filter joins adjacent regions by closing. After applying the erosion filter, unconnected regions return to their original state, while connected regions stay connected for the most part.

In Figure 7, this filter also solves the problem of small holes appearing in some segmentation regions. This problem was most commonly seen in neural networks, where initial candidate selection was left out of the segmentation voxels that were largely surrounded by others chosen as candidates.
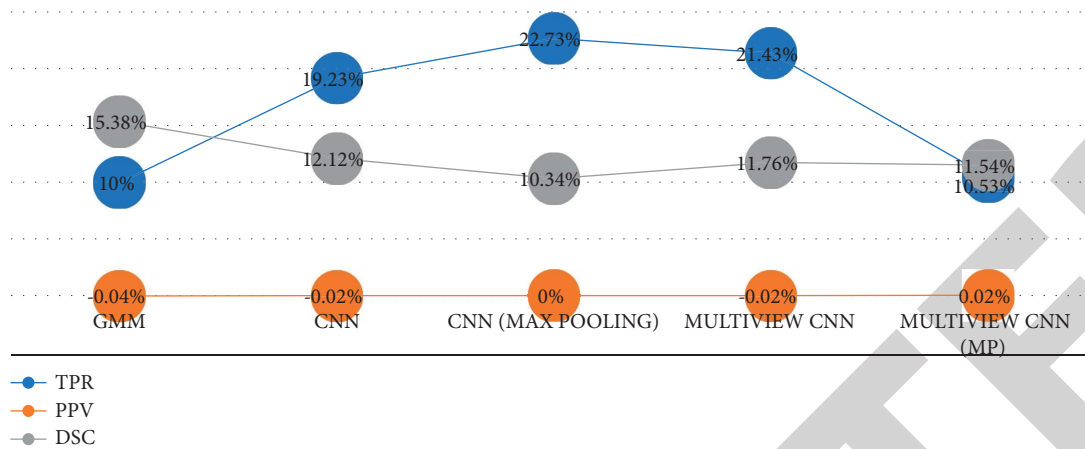
FIGURE 7: Changes in the metrics after the application of postprocessing.

FIGURE 8: Example of the benefits of the closing filter.

In Figure 8, It should be noted that this filter increases the number of false positives in the segmentation since it allows those regions corresponding to false positives to connect in a similar way to lesions. However, given the improvement observed in the TPR and DSC metrics, it can be said that the advantages of applying the filter outweigh its disadvantages, especially since the filter attacks the weaknesses observed in the techniques, which were able to obtain a segmentation with a low percentage of both false positives and hits to lesions.

5. Conclusions

For the MS Lesion Segmentation training database, two alternative methodologies of the state of the art were implemented, along with minor changes. Based on outlier detection using GMM and segmentation using convolutional neural networks, these algorithms had not been tested in this database and had no performance records. Both approaches achieve precise segmentation of MS lesions with good PPV. The fundamental issue with both approaches was the poor average TPR. Convolutional neural networks performed the best, successfully classifying multiple sclerosis lesions from the first set of candidates. It is worth noting that the number of photos used to create these segmentations was rather small, around 1.5% of all possible patches. In this regard, the poor TPR seen due to the small number of

patches utilized and the issues observed in automatic matter segmentation may be improved by changing the patch selection for this technique. However, segmentation produced by detecting GMM outliers has several advantages over CNN. CNN's estimate is partially based on the initial injury candidate selection. Using solely white matter voxel-cantered patches presents issues if the white matter segmentation has faults due to picture abnormalities. The GMM could not detect this issue because these abnormalities tend to show as outliers. Due to the minimal number of parameters in the GMM, the EM algorithm executes quickly, allowing for rapid segmentation of lesions. Using the GPU reduces the training time of CNNs to a level comparable to the time necessary to acquire a segmentation using the GMM in the CPU. Finally, applying a closure filter to the segmentation findings improved them significantly, alleviating underestimating the lesion size reported in both techniques.

6. Recommendation

Using the suggested methodologies, the segmentation quality can be improved in the state of the art for the database. A contiguous region for a single lesion is not always attainable because the approach does not employ information about the voxel's immediate surroundings, which causes segmentation issues. Instead, the GMM might

include voxel-specific properties. A voxel's near surroundings, as well as intensities from each imaging modality, may help segmentation. CNN can be improved greatly. Subsampling candidate voxels removes lesions not used in the model's training. Using different MRI modalities to identify outliers could improve the candidate voxel sampling step. For generalization, neural networks require a lot of data. The networks' outcomes may improve if they access more images from varied patients. The neural networks work for MRI at the same resolution they were trained on. Adding more data preprocessing stages would allow input images to match the training images in resolution.

Data Availability

The data underlying the results presented in the study are available within the manuscript.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] M. Abadi, P. Barham, J. Chen et al., "TensorFlow: a system for large-scale machine learning," in *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation*, Savannah, GA, USA, November 2016.

[2] S. Lindgren Belal, M. Sadik, R. Kaboteh et al., "Deep learning for segmentation of 49 selected bones in CT scans: first step in automated PET/CT-based 3D quantification of skeletal metastases," *European Journal of Radiology*, vol. 113, pp. 89–95, 2019.

[3] A. R. Cohen, "Boston children's hospital and the origin of pediatric neurosurgery," *Child's Nervous System*, vol. 30, no. 10, pp. 1621–1624, 2014.

[4] M. L. Thivagar, A. S. Al-Obeidi, B. Tamilarasan, and A. A. Hamad, "Dynamic analysis and projective synchronization of a new 4D system," in *IoT and Analytics for Sensor Networks*, pp. 323–332, Springer, Singapore, 2022.

[5] V. Hemanth, V. S. Rajinikanth, S. Rao, N. M. S. Mishra, R. Vijayarajan, and S. Arunmozhi, "Image fusion practice to improve the ischemic-stroke-lesion detection for efficient clinical decision making," *Evolutionary Intelligence*, vol. 14, no. 2, pp. 1089–1099, 2021.

[6] M. Filippi, P. Preziosa, B. L. Banwell et al., "Assessment of lesions on magnetic resonance imaging in multiple sclerosis: practical guidelines," *Brain*, vol. 142, no. 7, pp. 1858–1875, 2019.

[7] A. A. Hamad, M. L. Thivagar, M. B. Alazzam et al., "Dynamic systems enhanced by electronic circuits on 7D," *Advances in Materials Science and Engineering*, vol. 2021, Article ID 8148772, 11 pages, 2021.

[8] P. Golland, N. Hata, C. Barillot, J. Hornegger, and R. Howe, "Medical image computing and computer-assisted intervention," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, Boston, MA, USA, September 2014.

[9] M. A. Guvakova, "Improving patient classification and biomarker assessment using Gaussian Mixture Models and Bayes' rule," *Oncoscience*, vol. 6, no. 11-12, pp. 383–385, 2019.

[10] S.-W. Jang and M. Jung, "Exposed body component-based harmful image detection in ubiquitous sensor data," *Cluster Computing*, vol. 22, no. S1, pp. 1899–1910, 2019.

[11] G. Alshammari, A. A. Hamad, Z. M. Abdullah et al., "Applications of deep learning on topographic images to improve the diagnosis for dynamic systems and unconstrained optimization," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 4672688, 7 pages, 2021.

[12] C. J. Karakasis, A. O. Bricker, and S. E. Jones, "Magnetic resonance imaging in multiple sclerosis," *Multiple Sclerosis and Related Disorders*, vol. 8, no. 5, Article ID a028969, 2018.

[13] A. Karimian and S. Jafari, "A new method to segment the multiple sclerosis lesions on brain magnetic resonance images," *Journal of Medical Signals & Sensors*, vol. 5, no. 4, pp. 238–244, 2015.

[14] H. Lee and J. Song, "Introduction to convolutional neural network using Keras; an understanding from a statistician," *Communications for Statistical Applications and Methods*, vol. 26, no. 6, pp. 591–610, 2019.

[15] Y. Li and H. Wu, "A clustering method based on K-means algorithm," *Physics Procedia*, vol. 25, pp. 1104–1109, 2012.

[16] J. Zhao, D. Li, J. Pu, Y. Meng, A. Sbeih, and A. A. Hamad, "Human-computer interaction for augmentative communication using a visual feedback system," *Computers & Electrical Engineering*, vol. 100, Article ID 107874, 2022.

[17] S. B. Nemade and S. P. Sonavane, "Image segmentation using convolutional neural network for image annotation," in *Proceedings of the International Conference on Communication and Electronics Systems (ICCES)*, pp. 838–843, Coimbatore, India, July 2019.

[18] A. Shakourirad and M. A. Md, "Magnetic resonance imaging abnormalities in multiple sclerosis," *A Review*, vol. 116, no. 1, pp. 135–146, 2015.

[19] A. A. Hamad, M. L. Thivagar, and J. Alshudukhi, "Talal saad alharbi, saud aljaloud, khalid twarish alhamazani, zelalem meraf, secure complex systems: a dynamic model in the synchronization," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 9719413, 6 pages, 2021.

[20] C. G. B. Yogananda, B. C. Wagner, G. K. Murugesan, A. Madhuranthakam, and J. A. Maldjian, "A deep learning pipeline for automatic skull stripping and brain segmentation," in *Proceedings of the IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 727–731, Venice, Italy, April 2019.

[21] J. Zambrano, "Gaussian Mixture Model - method and application," *FUture DIrections for Process industry Optimization*, vol. 178, 2017.