

Received March 26, 2021, accepted April 26, 2021, date of publication May 3, 2021, date of current version May 11, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3077006

# Multi-Class Probabilistic Atlas-Based Whole Heart Segmentation Method in Cardiac CT and MRI

TARUN KANTI GHOSH<sup>1</sup>, MD. KAMRUL HASAN<sup>1</sup>,  
SHIDHARTHO ROY<sup>2</sup>, (Student Member, IEEE), MD. ASHRAFUL ALAM<sup>2</sup>,  
EKLAS HOSSAIN<sup>3</sup>, (Senior Member, IEEE), AND MOHIUDDIN AHMAD<sup>2</sup>, (Member, IEEE)

<sup>1</sup>Department of Biomedical Engineering, Khulna University of Engineering and Technology, Khulna 9203, Bangladesh

<sup>2</sup>Department of Electrical and Electronic Engineering, Khulna University of Engineering and Technology, Khulna 9203, Bangladesh

<sup>3</sup>Oregon Renewable Energy Center (OREC), Department of Electrical Engineering and Renewable Energy, Oregon Institute of Technology, Klamath Falls, OR 97601, USA

Corresponding author: Md. Kamrul Hasan (m.k.hasan@eee.kuet.ac.bd)

**ABSTRACT** Accurate and robust whole heart substructure segmentation is crucial in developing clinical applications, such as computer-aided diagnosis and computer-aided surgery. However, the segmentation of different heart substructures is challenging because of inadequate edge or boundary information, the complexity of the background and texture, and the diversity in different substructures' sizes and shapes. This article proposes a framework for multi-class whole heart segmentation employing non-rigid registration-based probabilistic atlas incorporating the Bayesian framework. We also propose a non-rigid registration pipeline utilizing a multi-resolution strategy for obtaining the highest attainable mutual information between the moving and fixed images. We further incorporate non-rigid registration into the expectation-maximization algorithm and implement different deep convolutional neural network-based encoder-decoder networks for ablation studies. All the extensive experiments are conducted utilizing the publicly available dataset for the whole heart segmentation containing 20 MRI and 20 CT cardiac images. The proposed approach exhibits an encouraging achievement, yielding a mean volume overlapping error of 14.5 % for CT scans exceeding the state-of-the-art results by a margin of 1.3 % in terms of the same metric. As the proposed approach provides better results to delineate the different substructures of the heart, it can be a medical diagnostic aiding tool for helping experts with quicker and more accurate results.

**INDEX TERMS** Bayesian framework, deep convolutional neural network, non-rigid registration, probabilistic atlas, whole heart segmentation.

## I. INTRODUCTION

In this section, we first specify the challenges and motivations of this article in subsection I-A. Secondly, we review several recent literature in subsection I-B for the particularized difficulties in subsection I-A. Finally, in subsection I-C, we summarize our contributions in this article.

### A. PROBLEM PRESENTATION

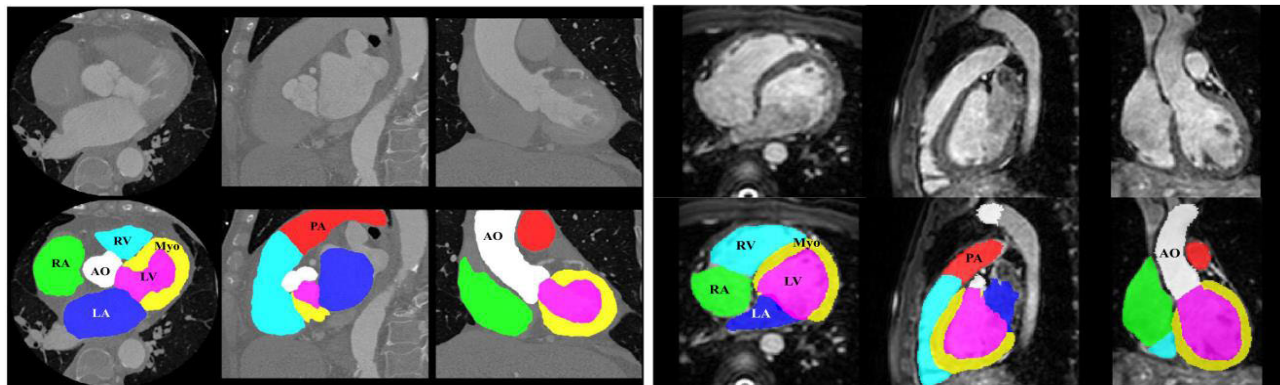
Medical imaging and computing technologies have revolutionized modern medicine and healthcare, which are increasingly crucial for the treatments and diagnosis of different Cardiovascular Diseases (CVDs) [1], [2]. Currently, various

non-invasive cardiac imaging assessments, such as Magnetic Resonance Imaging (MRI), Computed Tomography (CT), Ultrasound (US), Positron Emission Tomography (PET), and Single Photon Emission Computed Tomography (SPECT), are being used widely for clinical and diagnostic applications in cardiology [3], [4]. Among all of those imaging modalities, the MRI and CT scans have essential functions in the non-invasive evaluation of CVDs through extensive experimentation, and clinical applications [1], [5]. The CT images are more commonly employed than MRI images due to their high-speed retrieval and more affordable expense. In contradiction, the MRI images have no ionizing radioactivity and outstanding soft-tissue contrast. However, most of the current clinically convenient image examination approaches either attunes for the MRI or CT images alone [1], [5].

The associate editor coordinating the review of this manuscript and approving it for publication was Mouloud Denai<sup>1</sup>.

**TABLE 1.** Details of different heart substructures to be segmented from the whole chest images and their anatomical locations.

Substructures	Acronym	Anatomical position and functions
Left Ventricular cavity	LV	The bottom left portion of the heart below the left atrium for pumping oxygenated blood to all tissues
Right Ventricular cavity	RV	The bottom right portion of the heart below the right atrium for pumping oxygen-depleted blood to the lungs
Left Atrial cavity	LA	The upper left portion of the heart to receive oxygenated blood from the four pulmonary veins
Right Atrial cavity	RA	The upper right portion of the heart for returning deoxygenated blood from the body to the RV
Myocardium of LV	Myo	The myocardium is the muscular middle layer of the wall of the heart for pumping blood around the body
Ascending Aorta	AO	The ascending aorta is connected to the heart's LV to allow the flow of blood from the heart into the aorta
Pulmonary Artery	PA	The pulmonary artery begins at the base of the heart's RV to deliver oxygen-depleted blood to each similar lung



**FIGURE 1.** Illustration of complexities for achieving robust WHS in the employed cardiac MM-WHS-2017 dataset with corresponding ground-truth, borrowed from [1], where the left three columns display the three orthogonal views of a cardiac CT image and its corresponding WHS result, and the right three columns exhibit example cardiac MRI data and the WHS result. The LV, RV, LA, RA, Myo, AO, PA, and background in the first columns have a different shape, texture, and spatial location than the other two CT columns (second and third). Similar complexities are notified for the MRI scans (last three columns).

Delineating different vital volumetric substructures from the whole volumetric medical images are of great significance for clinical practice for quantifying the morphological and pathological changes [1]. Precise segmentation aids the subsequent quantitative evaluation of the Volume of Interests (VOIs). It also avails specific diagnosis, forecast of prognosis, computer-aided diagnosis, radiation therapy, computer-aided surgery, intra-operative guidance, and surgical planning [1], [6]. For instance, liver segmentation from 3D abdominal CT scans is a crucial prerequisite for tumor resection, computer-assisted live donor transplantation, and minimally invasive surgery interventions [7]–[9]. Besides, for the treatment of CVDs, including radio-frequency ablation and surgical preparation of crucial congenital heart diseases, volumetric cardiac MR image segmentation is necessary [10]–[12].

Currently, the Whole Heart Segmentation (WHS) is an imperative preliminary action for a wide range of clinical treatments. For example, the pathology localization and accurate ventricular dimensions [13], [14], which aims to delineate seven different heart substructures, as outlined in Table 1, from the whole cardiac images (see in Fig. 1). The precise heart quantification requires the subtle segregation of different heart substructures. For instance, the ejection portion and the myocardial mass are estimated from the segmented ventricular and the myocardial results, respectively, which are critical pointers of cardiac disorder detection. The whole heart's manual delineation is labor-intensive and

tedious, necessitating almost eight hours for an individual subject [15]. Consequently, the designing of computer-aided methods to investigate medical images automatically are highly demanding. However, an automated segmenting of a whole heart is also challenging due to the cardiac anatomical shape variations and the indefinite borders among different heart substructures [1], as depicted in Fig. 1. Achieving entirely computerized WHS is arduous due to the following hurdles:

- Wide shape, structure, and boundary variations of the cardiac anatomy and their subsequent cardiac imaging.
- Indefinite boundaries, with inadequate edge information, between different heart substructures in cardiac images.
- Low cardiac image quality and seldom they blend with other visually alike substructures (see in Fig. 1).

### B. RECENT METHODS

Convolution Neural Network (CNN)-based Deep Learning (DL) strategies are the most commonly employed medical image segmentation method in various application fields and modalities [16]–[21]. Table 2 bestows various DL-based and other techniques for the WHS of different substructures, including their corresponding utilized datasets and identical outcomes. UNet, initially proposed in [31], is mostly applied CNN structure for the WHS [4], [18]–[20], [27], [29], [32]–[34], which can be summarized as follows:

**TABLE 2.** Several published literature for WHS with their employed datasets and achievements conferring varying metrics, such as mDSC, mVOE, and mSn, respectively, for mean dice similarity coefficient, mean volume overlapping error, and mean sensitivity.

Different methods	Year	Datasets	Metrics		
			mDSC	mVOE	mSn
A multi-modality atlas with a distinct label merging method based on a multi-scale patch approach and a unique global atlas ranking system [15]	2016	MM-WHS-2017	0.899	0.183	–
An automated approach using dilated convolutional neural networks for aggregating features at various scales through convolutional layers with too fewer parameters [22]	2016	HVSMR-2016	0.865	0.238	–
A multi-planar deep CNNs with an adaptive merging procedure utilizing corresponding information from the separate planes of the 3D scans for enhanced delineations [5]	2017	MM-WHS-2017	0.851	0.259	0.831
An atlas-based segmentation approach with a two-stage registration pipeline, where the majority voting and STEPS algorithms were used for the merging of the labels [23]	2017	HVSMR-2016	0.900	0.182	–
A method having three crucial actions: heart identification from landmark detection, heart isolation using mathematical shape model, and segmentation using learning-based voxel classification and local phase analysis [24]	2018	HVSMR-2016	0.760	0.387	–
A deeply-supervised fractal network, where the multi-paths with different receptive fields were established in a self-similar fractal system to obtain the hierarchical features [25]	2018	HVSMR-2016	0.760	0.387	–
Two-stage UNet framework, where the first stage detected the VOIs and the second stage accurately segmented the heart substructures [4]	2019	MM-WHS-2017	0.793	0.342	–
A semi-supervised approach, where the student model acquires from labeled target data and also searches unlabeled target data and labeled data by two teacher models [26]	2019	MM-WHS-2017	0.860	0.245	–
Category attention boosting module connecting the deep network estimation graph with the boosting approach [20]	2019	HVSMR-2016	0.894	0.191	–
A CNN-based architecture, which includes principal component analysis as a supplementary data enlargement routine [27]	2019	MM-WHS-2017	0.890	0.198	–
A deep heterogeneous feature gathering network (HFANet) to wholly employ corresponding knowledge from various views of 3D cardiac data [28]	2019	HVSMR-2016	0.942	0.110	–
	2019	MM-WHS-2017	0.909	0.167	–
	2019	AAPM-2017	0.883	0.209	–
An adversarial training strategy for training the networks using the UNet as a generator in adversarial network [29]	2020	HVMSR-2016	0.870	0.230	0.890
Cascading of two volumetric FCNs, where the first network aimed at locating the cardiac area, during the second segmented different cardiac and great vessel substructures [21]	2020	HVSMR-2016	0.942	0.109	0.943
A pipeline of two FCNs, where the first network localizes the bounding box's center encompassing all heart substructures, while the second network segments them [30]	2020	MM-WHS-2017	0.908	0.168	–

The authors in [27] suggested a framework consisting of two 3D-UNets, where the first network was employed to localize the bounding box encompassing the heart, and the second network was used for the fine segmentation of different substructures. They also employed principal component analysis-based image augmentation for enhancing the WHS results. An UNet-based Omega-Net was introduced in [18] consisting of a set of UNet for fine-grained WHS. By explicitly finding the VOIs and turning the input image to a standard orientation, this method adequately accommodated the loss of segmentation precision produced by distinguishing between training and testing images. Faster R-CNN (FRCNN) and UNet were combined in [19] naming as a CFUN. With the FRCNN's precise localization ability and UNet's strong segmentation capability, CFUN needed solely one-step detection and segmentation deduction to receive the VOIs. The authors also used a novel loss function based on border information to accelerate the training and enhance the WHS results. The author in [32] applied an approach combining with shape context knowledge encoded in volumetric shape models consisting of three primary steps: explorer segmentation with orthogonal 2D-UNet, shape context estimation, and refining segmentation with UNet and shapes context. The additional shape information, also called shape context, was applied for implementing explicit 3D shape knowledge to CNN. The authors in [33] developed a pipeline

with three quintessential steps: employment of 3D-UNet for coarse detection of VOIs to alleviate surrounding tissues' impact, artificially augmentation of the training dataset by extracting different VOIs, and a refined 3D-UNet for segmentation refinement using the augmented training dataset. The authors in [20] joined the attention tool into the gradient expanding process for enhancing the coarse segmentation information with less computation expense. Moreover, they introduced the Category Attention Boosting (CAB) module into the 3D-UNet network and constructed a new multi-scale boosting model, CAB-UNet, extending the network's gradient flow and making sole usage of the low-resolution feature information.

However, the segmentation strategies relying on the UNet-based structure suffer from redundant fusion tactics, imposing concatenation only at the identical scale feature maps of the encoder and decoder [16], [17], [35]. There are some other DL- and Atlas-based techniques for the WHS [5], [15], [21], [25], [26], [28], [30], [36], which are also reviewed and summarized as follows:

A deep poincaré map encapsulating prior knowledge with a dynamical system was applied in [36]. A CNN-based approach had then navigated an ambassador over the cardiac MRI image, moving to approach a route that outlined the target substructures. Different CT or MRI plans, such as coronal, sagittal, and axial, are explored for an adaptive

fusion in [5] with a multi-planar deep CNN-based method. The authors independently built and trained three CNNs, with the same constructive arrangement, for those three planes and finally combined them to obtain the segmented VOIs. A pipeline having two Fully Convolutional Networks (FCNs) was implemented in [30], where the first CNN localized the bounding box's center around the different substructures and the succeeding second CNN concentrated on these regions for segmenting the organs. The authors in [28] developed a deep Heterogeneous Feature Aggregation Network (HFANet) for entirely exploiting corresponding information from 3D cardiac data. They utilized asymmetrical 3D kernels and pooling for obtaining heterogeneous features in identical encoding routes. Therefore, distinguishable features were extracted from a specific view, and necessary contextual information was kept. Then, they employed a content-aware multi-planar concatenation for aggregating meaningful features to boost the WHS performance. Further, to overcome the model size, they also devised a new DenseVoxNet model by sparsifying skip connections trained in an end-to-end manner. A Dual-Teacher strategy was suggested in [26], where the student model acquired precisely the knowledge of unlabeled target data from intra-domain teachers by fostering prediction texture and the shape priors embedded in labeled source data from inter-domain teachers via information distillation. The authors also examined the utility of concurrently leveraging unlabeled data and well-known cross-modality data for the segmentation. A 3D fractal network for effective computerized segmentation technique was introduced in [25]. The designed network took complete convolutional construction to implement effective, well-defined, and volume-to-volume prognostication. Prominently, by recursively employing a single augmentation rule, the authors assembled the network in a novel self-similar fractal system and consequently promoted it in consolidating hierarchical evidence for precise segmentation. They also employed a deep supervision tool to mitigate the vanishing gradients obstacle and increase our network's training effectiveness on inadequate medical image datasets. The authors in [15] offered a multi-scale patch for hierarchical local atlas ranking. Their segmentation approach exercised multi-modality atlases from MRI and CT and embraced a new label merging method based on the recommended multi-scale patch policy and a new global atlas ranking scheme. Both the local and global atlas ranking actions used the information-theoretic criteria to estimate the association between the target image and the atlases from versatile modalities.

### C. OUR CONTRIBUTION

While many approaches have already been developed and implemented for the WHS, there is still room for performance improvement. This article proposes a statistical WHS method that joins the prior anatomical information described by probabilistic atlas into the Bayesian inference for delineating seven different heart substructures (see in subsection I-A) in

cardiac CT and MRI images. This article aims to contribute the following items for the WHS:

- A multi-class WHS framework based on the proposed non-rigid registration pipeline for atlas construction (see in subsection II-B 1), utilizing a multi-resolution strategy for obtaining the highest possible mutual information between moving and fixed CT or MRI images.
- Fine-tuning different registration algorithm parameters, such as optimizer, interpolator, metric, resampling technique, and transformation, to achieve better spatial alignment between moving and fixed CT or MRI images.
- Development of various fusion strategies of multiple atlases to delineate the different anatomical structures for ablation studies.
- Incorporation of proposed non-rigid registration pipeline into the statistical Bayesian inference and expectation-maximization for the WHS task.
- Implementation of other CNN-based supervised methods, where we employ three variants of encoder-decoder networks to compare them with the proposed statistical framework.
- Validation of all the extensive experiments utilizing the same publicly available dataset named MM-WHS-2017 (see details in subsection II-A), providing the state-of-the-art results for the WHS to our most trustworthy knowledge.

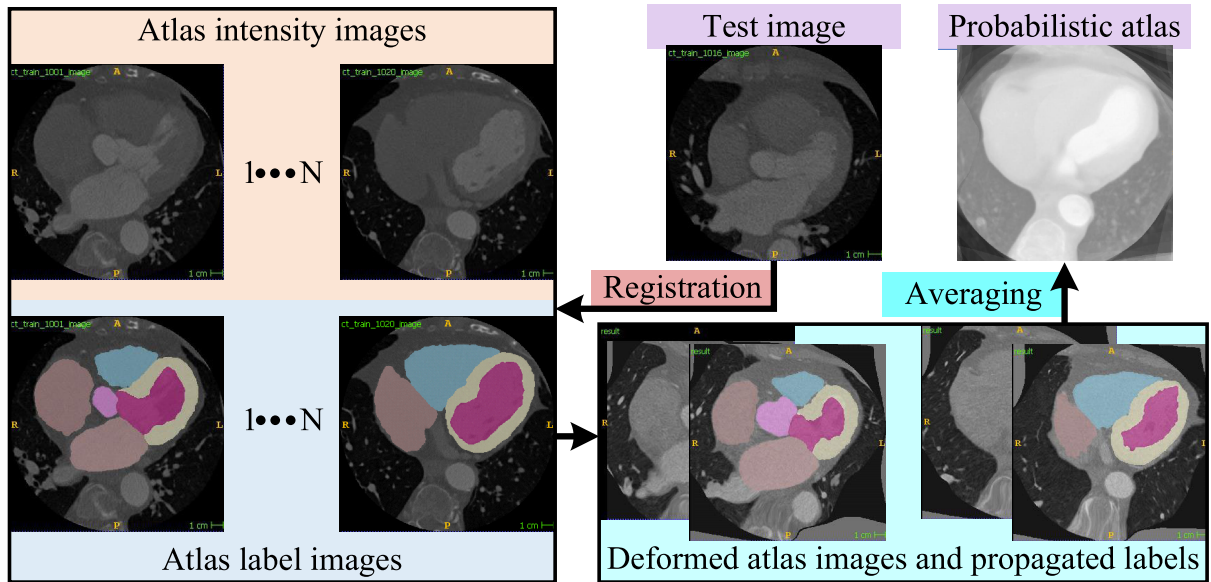
The rest of the sections are manifested as follows: section II illustrates the used datasets to bear extensive experiments and the different methodologies. Section III describes the obtained results accompanying a precise analysis and state-of-the-art connections. Finally, section IV terminates the paper with prospective future acts.

## II. MATERIALS AND METHODS

This section elaborates on the materials and methodologies in the article. We present the utilized dataset and our proposed WHS schemes in subsections II-A and II-B, respectively. Subsections II-C and II-D respectively explain other implemented CNN-based methods for WHS and the hardware & metrics used to evaluate the experimentation.

### A. DATASET

All the comprehensive experiments were conducted utilizing publicly available MM-WHS-2017 dataset [1], as it is commonly used in recent articles (see in Table 2), which contains 40 cardiac whole heart volumetric MRI and CT data. The VOIs include the seven different substructures in the utilized WHS dataset, as described earlier in subsection I-A (see in Table 1). The different substructures, such as LV, RV, LA, RA, Myo, AO, and PA (see in Fig. 1) of both the CT and MRI are labeled as 500, 600, 420, 550, 205, 820, and 850, respectively. We aim to segment those seven organs from both the cardiac CT and MRI scans. Hence, we have termed it a multi-class (8-classes) segmentation task,



**FIGURE 2.** Illustration of probabilistic atlas-based segmentation for the whole heart segmentation employing different fusion strategies, where we use non-rigid registration for atlas construction.

including the background and seven different heart substructures. All the experimental WHS approaches are assessed following a leave-one-out evaluation strategy [37].

The 20 CT and 20 MRI sequences were collected from a 64-slice CT scanner (Philips Medical Systems, The Netherlands) and a 1.5T clinical scanner (Philips Healthcare, The Netherlands), respectively. The volumes from different scanners were stored as NIfTI file format in differing image properties and resolutions to provide training data to promote more robust algorithms’ construction [38]. The former volumetric CT images were obtained in axial view, incorporating the entire heart from the topmost abdominal to the aortic arch with an in-plane resolution of  $0.44 \times 0.44$  mm, and the standard slice thickness of 0.60 mm. The latter volumetric MRI images were accumulated with a resolution of  $2 \times 2 \times 2$  mm and reconstructed to nearby  $1 \times 1 \times 1$  mm.

### B. PROPOSED METHOD

The WHS’s proposed method essentially consists of two integral parts, such as *Atlas construction* and *segmentation strategy*, where we integrate several approaches to determine the best performer for WHS. The elaborate discussion of these two crucial parts of our proposed framework is manifested in the following two subsections, II-B1 and II-B2, respectively.

#### 1) REGISTRATION AND ATLAS CONSTRUCTION

Fig. 2 demonstrates the proposed pipeline for atlas-based WHS, where the registration is the crucial integral step to deform a moving CT or MRI image to align with a fixed CT or MRI image spatially. Let us consider that  $I_T$  is an unseen fixed cardiac image (CT or MRI) to be segmented ( $L_T$ ).  $\{(I_i, L_i)|i = 1, \dots, N\}$  is a set of moving images to build

atlases, where  $I_i$ ,  $L_i$ , and  $N$  are the  $i^{th}$  intensity image, its corresponding  $i^{th}$  label image, and the number of member images in the reference volumes, respectively. We perform moving-to-fixed registration (non-rigid) to construct a deformed set,  $\{(I_i^D, L_i^D)|i = 1, \dots, N\}$ , where  $I_i^D$  and  $L_i^D$  are the resultant deformed intensity image and its corresponding deformed label image. The deformed intensity image ( $I_i^D$ ) and its corresponding deformed label image ( $L_i^D$ ) are averaged to construct probabilistic atlas ( $I_A$ ) and probabilistic label ( $L_A$ ), respectively, using (1) and (2), respectively.

$$I_A(x, y, z) = \frac{1}{N} \sum_{i=1}^N I_i^D(x, y, z) = \frac{1}{N} \sum_{i=1}^N I_i(T_{NR_i}(x, y, z)), \quad (1)$$

$$L_A(x, y, z) = \frac{1}{N} \sum_{i=1}^N L_i^D(x, y, z) = \frac{1}{N} \sum_{i=1}^N L_i(T_{NR_i}(x, y, z)), \quad (2)$$

where  $T_{NR_i}$  is a transformation between  $i^{th}$  pair of  $I_i$  and  $I_T$ . For  $M$  voxels and  $k \in \{1, 2, \dots, K = 7\}$  target substructure regions, both the  $I_A(x, y, z)$  and  $L_A(x, y, z)$  probabilistic atlas is a matrix with  $N \times K$  elements, and each element ( $P_{nk} \in (I_A \text{ or } L_A)$ ) represents the anatomical knowledge about the heart provided by training samples on the prior probability of  $n^{th}$  voxel belonging to a particular tissue class  $k$ .

The intensity image ( $I_i$ ) is appointed as the moving image for deforming to the unseen fixed images ( $I_T$ ) using a non-rigid transformation ( $T_{NR}$ ) [39]–[41]. The algorithm of

$T_{NR}$  is a combination of a global and local transformations as  $T_{NR}(x, y, z) = T_{global}(x, y, z) + T_{local}(x, y, z)$ . The global ( $T_{global}$ ) is an affine transformation allowing scaling, translation, rotation, and shearing of  $I_i$ , whereas the local ( $T_{local}$ ) is a free-form deformation model based on B-splines [42]. However, in our pipeline, we formulate the registration as an optimization problem in (3) for maximizing the similarity between  $I_i$  and  $I_T$  employing the  $T_{NR}$ .

$$\overline{T_{NR}} = \underset{T_{NR}}{\operatorname{argmin}} \Upsilon[T_{NR}; I_T(x, y, z), I_i(x, y, z)], \quad (3)$$

where  $\overline{T_{NR}}$  is an optimal transformation to spatially align  $I_i(T_{NR}(x, y, z))$  to  $I_T$ .  $\Upsilon$  is a cost function, which we minimize by designing a pipeline combining different algorithms. The cost function ( $\Upsilon$ ) is optimized using adaptive stochastic gradient descent optimizer [43]. Other crucial algorithms in our proposed pipeline, such as interpolator, metric, and resample-interpolator, are the B-Spline algorithm with an order of 1, mutual information [44] with histogram bins of 32, and B-Spline algorithm with an order of 3, respectively. A multi-resolution strategy, with resolutions of 4, is used to bypass local minima [45]. We use *Elastix-5.0*<sup>1</sup> [46] to implement our proposed registration pipeline. In this work, a complete probabilistic atlas is built and evaluated following a leave-one-out evaluation strategy [37].

## 2) SEGMENTATION STRATEGIES

This subsection presents and designs several integral strategies in our proposed pipeline (see in Fig. 2) to delineate different anatomical substructures (see in Table 1); once the moving images are deformed to a fixed image employing a non-rigid registration.

### a: MULTI-ATLAS LABEL PROPAGATION

The easiest and quickest technique to assign a label to each voxel of an input test image is the label propagation of deformed labels ( $L_i^D$ ) to the unseen test image space [47]–[49]. Multi-atlas Label Propagation (MALP) trades better with the registration errors comparing a single atlas and better accounts for anatomical variability [50]. The MALP also strengthens over single label propagation, as it can reject outliers (minority) of the deformed labels ( $L_i^D$ ). The Majority Voting Fusion (MVF) [15] is a MALP approach, which counts the number of atlases provided the same label for a test voxel,  $n$ , in the exact spatial location. The resultant segmented VOI ( $L_T$ ) can be estimated from  $L_i^D$ , as in (4):

$$L_T(n) = \operatorname{argmax}_{k \in \{1, \dots, K\}} \sum_{i=1}^N \Gamma(L_i^D(n), k), \quad (4)$$

where  $\{1, 2, \dots, K\}$  is a set of  $K$  ( $=7$ ) labels of the heart anatomy (see in subsection II-A),  $\Gamma(L_i^D(n), k)$  is a counting function, as defined in (5).

$$\Gamma(L_i^D(n), k) = \begin{cases} 1 & \text{if } L_i^D(n) = k, \\ 0 & \text{if } L_i^D(n) \neq k. \end{cases} \quad (5)$$

The other method of MALP is the estimation median value of all the candidate labels in a particular voxel location. The median value of each test voxel ( $n$ ) from the deformed labels ( $L_i^D$ ) also provides a robust estimation of resultant label ( $L_T$ ) [51], which can be formulated as  $L_T(n) = \operatorname{Median}(L_k^S(n))$ ,  $\forall k \in K$ , where  $L_k^S(n)$ ,  $\forall k \in K$ , is a stacked (in  $4^{\text{th}}$  dimension) deformed label of  $L_i^D$ ,  $\forall i \in N$ , which has  $N$ -values for all voxels,  $n$ .

### b: PROBABILISTIC ATLAS-BASED SEGMENTATION (PAS)

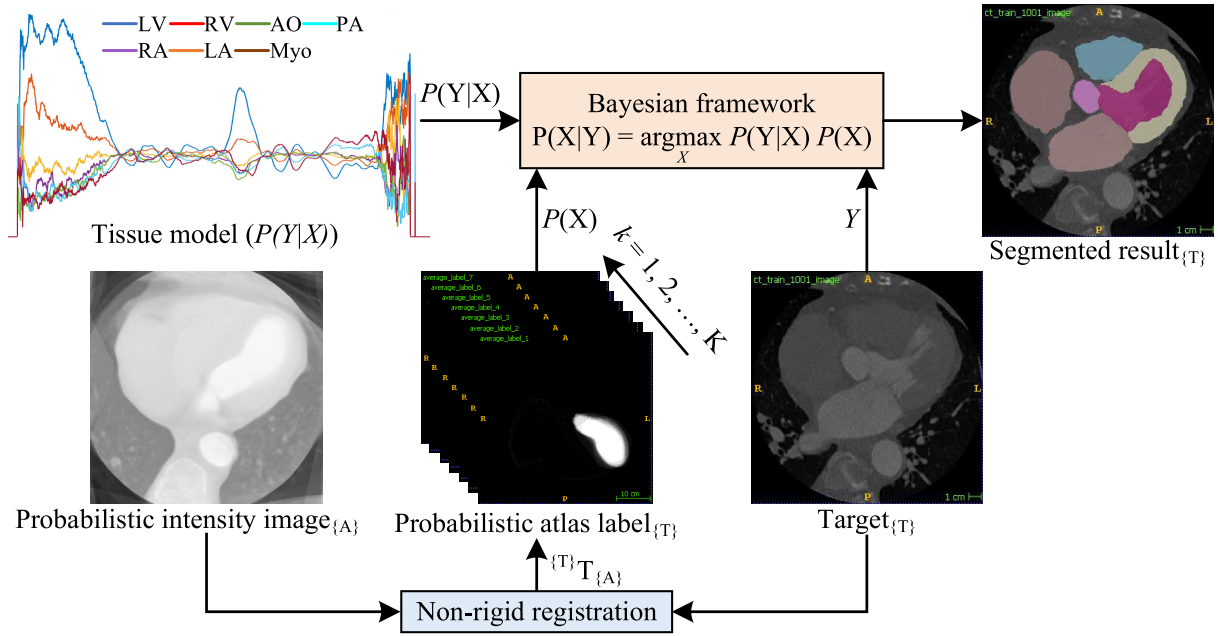
In our proposed PAS framework, we define the actual moving labels as  $X$  and the target fixed image as  $Y$ , where the components of  $X$  and  $Y$  are prepared by a spatial location expressed by  $n \in J$ , wherever  $J$  is the simplistic 3D rectangular grid index  $(x, y, z)$ . Let us considering that  $X = (x_1, x_2, \dots, x_M)$ ,  $Y = (y_1, y_2, \dots, y_M)$ , and  $A = (a_1, a_2, \dots, a_M)$  are the sample realizations of label image, intensity image, and probabilistic label, respectively, where  $M$  is the total voxel number. Example space of  $X$  is indicated by  $\Omega_x$ , where  $\Omega_x = \{x : x_n \in \{k = 1, 2, \dots, K\}, \forall n \in J\}$ . The probability atlas is  $K$ -vector  $a_j = (a_{j1}, a_{j2}, \dots, a_{jK})$ ,  $\forall j \in M$ , where all the ingredient corresponds to expectation of  $K$ -different heart substructures. The prior probability for all voxels can be expressed as  $P(x_n = k) = a_{nk}$ ,  $\forall n \in J; k \in K$ .

The hindrance comprises determining the label  $X$  that adequately illustrates the provided observation  $Y$  according to any loss function. As a decision rule, we chose MAP (maximum a posteriori) and the formula for the realization of estimating of  $X$  as  $\hat{x} = \operatorname{argmax}_x P(X = x | Y = y)$ . The posterior probability ( $P(X|Y)$ ) can be written as the multiplication of probability distribution ( $P(Y|X)$ ), also named tissue model [37], [52], and prior probability ( $P(X)$ ), according to the Bayes theorem.  $P(Y|X)$  is defined by signal intensity tissue models immediately formed from the scans and hand-operated segmentation of the data set. An intensity value's histogram is constructed for each heart substructure considering the given volumes' voxels, which belong to it, using hand-operated segmentation. In this work, we estimate the probability distribution  $P(Y|X)$  of the given image  $Y$  for the provided appropriate segmentation  $X$  from training data. Algorithm 1 shows detail process of estimating  $P(Y|X)$  using the number of bins ( $N_b$ ) for the histograms as 4096. Fig. 3 exhibits an illustration of the Bayesian voxel classification algorithm consolidating the application of the probabilistic atlas of the WHS. Besides, the probabilistic atlas gives the probability distribution  $P(X)$  once it has been mapped onto the target space employing the corresponding registration method utilized in its creation.

### c: EXPECTATION MAXIMIZATION (EM)

The EM algorithm is an iterative approach that classifies data points with the Gaussian Mixture Model (GMM), updates model parameters with the newly classified data, and classifies data points with the new parameters. In our implementation, we use EM to learn the parameters of the GMM in an unsupervised fashion for the segmentation of different heart

<sup>1</sup><https://elastix.lumc.nl/doxygen/index.html>



**FIGURE 3.** Voxel classification algorithm overview for the WHS, where the probabilistic atlas labels are transformed to the unseen test image space  $\{T\}$  utilizing the atlas’ anatomical image. The probabilistic atlas, also known as the tissue models, and the unseen objective image are given to the Bayesian interface as a prior probability  $P(X)$ , conditional probability  $P(Y|X)$ , and set of intensity values  $Y$ , respectively. The Bayesian interface estimates the posterior probability for the segmentation of given  $X$ , maximizing the  $P(X) \cdot P(Y|X)$ .

**Algorithm 1** Estimation of  $P(Y|X)$  of the Image  $Y$  for Given Segmentation  $X$

- 1: Initialize the counters as  $C_b^k = 0, \forall b \in N_b; k \in K$
- 2: **for** all training images **do**
- 3:   Rescale the image ( $[0 \sim N_b - 1]$ ) and extract VOI
- 4:   Accumulate  $N^k, \forall k \in K$  with corresponding pixel numbers.
- 5:   Compute histograms  $H^k, \forall k \in K$ , and accumulate in counters  $C_b^k += H^k, \forall k \in K$
- 6: **end for**
- 7: Normalize histograms  $C^k / N^k, \forall k \in K$
- 8: Scale the histograms as  $C_b^k = \frac{C_b^k}{\sum_{k=1}^K C_b^k}, \forall b \in N_b; k \in K$ , so that  $\sum_{k=1}^K C_b^k = 1, \forall b \in N_b$

substructures from the whole cardiac CT or MRI images. Let us assume that  $X = (x_1, x_2, \dots, x_M)$  is an unseen target CT or MRI image, where  $x_i, \forall i \in M$ , is  $d$ -dimensional voxel intensities and  $M$  is total voxel numbers. Suppose that  $x_i, \forall i \in M$ , is formed in an IID order from an underlying density of GMM model ( $P(X)$ ) with  $K$  ingredients, as in (6).

$$P(X|\Theta) = \sum_{k=1}^K \alpha_k P_k(X|Z_k, \theta_k), \quad (6)$$

where  $P_k(X|Z_k, \theta_k)$ ,  $\alpha_k$ , and  $Z_k$  are the  $k^{th}$  ingredient of GMM with parameters  $\theta_k$ , mix-up weights describing the probability that a randomly chosen  $X$  is formed by  $k^{th}$

element, and a vector of  $K$  pointer variables that are jointly independent, respectively, where  $\sum_{k=1}^K \alpha_k = 1$ . Therefore, the entire set of parameters for a GMM with  $K$  elements is  $\Theta = \{\alpha_1, \dots, \alpha_K, \theta_1, \dots, \theta_K\}$ . We can model  $P_k(X|Z_k, \theta_k)$  as Gaussian density function as in (7) with the parameters of  $\theta_k = \{\mu_k, \Sigma_k\}$ .

$$P_k(X|\theta_k) = \frac{1}{(2\pi)^{d/2} |\Sigma_k|^{1/2}} \cdot e^{(-\frac{1}{2}(X-\mu_k)^T \Sigma_k^{-1}(X-\mu_k))}, \quad (7)$$

where  $\forall k \in K$ . The parameters in the above equation, such as  $\mu_k$  and  $\Sigma_k$ , respectively denote the mean and co-variance or standard deviation of  $k^{th}$  Gaussian density function,  $P_k(X|\theta_k)$ . From the known GMM parameters and using the Bayes rule, the membership probabilities of a given observed vector ( $x_i \in X$ ) can be written, as in (8):

$$W_{ik} = \frac{\alpha_k \cdot P_k(x_i|\theta_k)}{\sum_{j=1}^K \alpha_j \cdot P_k(x_i|\theta_j)}, \quad \forall k \in K, \quad (8)$$

where  $\sum_{k=1}^K W_{ik} = 1$ . In our pipeline, we propose to initial parameters of the EM algorithm using the probabilistic atlas. The convergence of the EM algorithm is recognized by estimating the log-likelihood value, as in (10), after each iteration and stopping when it seems not to increase significantly from one iteration to the succeeding.

$$\begin{aligned} \log(l(\Theta)) &= \sum_{i=1}^M \log(P(X|\Theta)) \\ &= \sum_{i=1}^M \left( \log \left( \sum_{k=1}^K \alpha_k \cdot P_k(X|Z_k, \theta_k) \right) \right). \end{aligned} \quad (10)$$

The optimization of the EM algorithm in our pipeline essentially includes the following two steps to obtain optimal parameters of the GMM.

**Expectation-step (E-step):** It includes the following three steps. Step 1:

- 1) Denote the current parameter values (such as  $\alpha_k$ ,  $\mu_k$ , and  $\Sigma_k$ ,  $\forall k \in K$ ) as  $\Theta$ , where  $\sum_{k=1}^K \alpha_k = 1$ .
- 2) Compute membership probabilities ( $W_{ik}$ ) having the size of  $M \times K$  using the equation mentioned earlier in (8),  $\forall x_i \in X$ ,  $1 \leq i \leq M$ ,  $\forall k \in K$ , where  $\sum_{k=1}^K W_{ik} = 1$ .
- 3) Compute the log-likelihood ( $L$ ) using (10) for the current parameters ( $\Theta$ ), which is termed as  $L_{old}$ .

**Maximization-step (M-step):** The matrix of membership weights ( $W_{ik}$ ) is used to update the parameters  $\Theta$  using the following three steps. Step 1:

- 1) Update the mixture weights ( $\alpha_k$ ,  $\forall k \in K$ ) by using  $\alpha_k^{new} = \frac{1}{M} \sum_{i=1}^M W_{ik}$ , where  $\sum_{k=1}^K \alpha_k = 1$ .
- 2) Update parameters,  $\theta_k = \mu_k$ ,  $\Sigma_k$ ,  $\forall k \in K$ , by using the following equations,

$$\mu_k^{new} = \frac{1}{\hat{M}_k} \sum_{i=1}^M W_{ik} \cdot x_i, \quad \forall k \in K,$$

$$\Sigma_k^{new} = \frac{1}{\hat{M}_k} \sum_{i=1}^M W_{ik} \cdot (x_i - \mu_k^{new})(x_i - \mu_k^{new})^T,$$

where

$$\hat{M}_k = \sum_{i=1}^M W_{ik}, \quad \forall k \in K.$$

- 3) Compute the log-likelihood ( $L$ ) using (10) for the updated parameters ( $\Theta_{new}$ ), which is termed as  $L_{new}$ .

We run the EM algorithm until the convergence, which means we repeat until there is no significant change between  $L_{old}$  and  $L_{new}$ . Finally, the membership probabilities ( $W_{ik}$ ) are employed to delineate WHS's different substructures.

#### d: PAS+EM

In our proposed PAS+EM algorithm, we use posterior probabilities ( $P(X|Y)$ ) and membership probabilities ( $W_{ik}$ ) from PAS and EM algorithms, respectively, to delineate the different structures for WHS. For doing so, we use the following equation, as in (10), to obtain the probabilities for our PAS+EM algorithm.

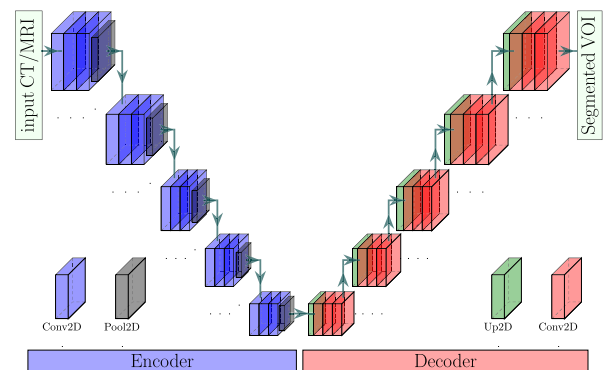
$$P_{PASEM} = W_{ik} \cdot P(X|Y), \quad \forall k \in K, \quad (10)$$

where all  $P_{PASEM}$ ,  $W_{ik}$ , and  $P(X|Y)$  are  $M \times K$  matrix. Finally, in the end, we apply  $P_{PASEM}$  to delineate the different structures for WHS.

### C. CNN-BASED METHODS

We have implemented a CNN-based semantic segmentation method besides our proposed atlas-based segmentation method to perform comprehensive ablation studies.

Segmentation of medical image has attained enormous progress notably since 2015 after the proposing of UNet [31]. Currently, CNN-based networks have been extensively practiced for the medical imaging domain, exceeding conventional image analysis techniques relying on hand-crafted features [53]. However, the CNN-based network for segmentation incorporates two fundamental elements: the encoder and the decoder [31]. An encoder consists of convolutional and pooling layers. The convolutional layers produce feature maps, whereas the pooling layers continuously decrease these feature maps' dimensions to gain more critical features with more eminent spatial invariance [17]. The decreased resolution feature maps also enlarge the maps' field-of-view and diminish the computational expense [54]. The decoder projects the distinctive lower resolution features discovered by the encoder onto the higher resolution pixel space to achieve a compact pixel-wise labeling [55]. However, the simple encoder-decoder network, named EDNet, in our implementation is depicted in Fig 4. The encoder in EDNet is a VGG-16 [56] with the pre-trained weight on ImageNet [57], which has five Convolutional Blocks (CB) and thirteen convolutional layers. Each CB's output is an input to the next CB through a pooling layer with a stride of  $2 \times 2$ . Hence, the encoder's output feature map has a resolution of  $m/2^5 \times n/2^5$  for an input resolution of  $m \times n$ . However, the decoder has five blocks to obtain the input resolutions of the output WHS masks ( $m \times n$ ), where we apply 2D upsampling, with a stride of  $2 \times 2$ , convolution with a kernel of  $3 \times 3$ , and a batch normalization [58] in each decoder block.

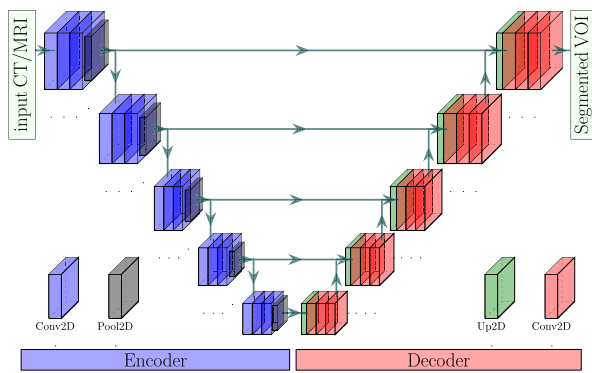


**FIGURE 4.** A pictorial presentation of EDNet applying a pre-trained VGG-16 network as an encoder for transferring the previous ImageNet knowledge to our WHS task. The exact number of pooling and upsampling layers are used in the encoder and decoder to regain the output's input resolutions.

However, the decreased feature maps due to pooling undergo spatial knowledge elimination injecting roughness, poor border knowledge, checkerboard artifacts, over-, and under-segmentation in the segmented substructures [17], [31], [54], [59]. To overcome these problems, the authors in [31] introduced skip connections in a UNet, permitting the decoder to retrieve the associated features discovered at

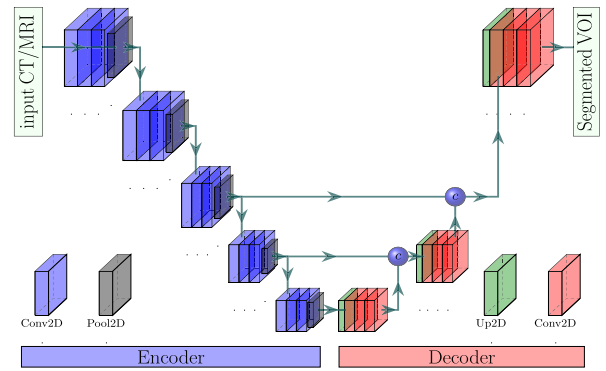


all encoder steps that were missed due to subsampling in the encoder. The feature maps from the encoder's antecedent layers are concatenated with the decoder's identical scale through the appliance of skip connections. Applying the skip connection of the popular UNet, we propose a VGG-UNet, where we have also employed the VGG-16 network as an encoder, as shown in Fig. 5. In our VGG-UNet, we apply the skip connections holding ladder-like compositions [60] motivated by UNet to succeed in the pooling weaknesses. All pooled layers of our network are concatenated channel-wise to a deconvoluted feature map with identical dimensions, where it acts as an offsetting link for the spatial knowledge dropped due to subsampling in the encoder.



**FIGURE 5.** The architecture of our VGG-UNet, where we concatenate the encoder's features with the same scale features of the decoder to compensate for the lost spatial information in the encoder. The exact number of pooling and upsampling layers are used in the encoder and decoder to regain the output's input resolutions.

Again, the authors in [54] combined features at varied coarseness levels of the encoder in their popular FCN to polish the output segmented VOIs. In this article, we propose a VGG-FCN, as shown in Fig. 6, with the pre-trained VGG-16 network in the encoder. The output feature map of such an encoder has 32-times fewer resolutions as VGG-16 has five pooling layers. The employment of  $32\times$  upsampling can output the segmented mask with the same size as the input image, called FCN32s. Such an upsampling produces a coarse or rough mask as it utilizes only global information from the more in-depth high-level features. The in-depth features are achieved while operating deeper, which causes the spatial location information lost. That indicates that output from shallower layers has more location knowledge. If we combine both local and global information, it can enhance the segmentation result. The output from the fifth pooling of VGG-16 is  $2\times$  upsampled and fused with the fourth pooling in our network. Then, the combined map is  $2\times$  upsampled and again linked with third pooling. Finally, we perform  $8\times$  upsampling, which provides a segmented WHS mask. Hence, the proposed VGG-FCN is also named VGG-FCN8s, where we use both the local and global information to obtain the final mask.



**FIGURE 6.** The structure of our VGG-FCN8s, where we apply a pre-trained VGG-16 as a feature learner. The output WHS masks are obtained from the fused feature maps from the fifth, fourth, and third pooling layers of VGG-16. The same number of pooling and upsampling layers are used in the encoder and decoder to regain the output's input resolutions.

We implemented three semantic segmentation network variants to compare them with the proposed Atlas-based segmentation method of WHS. We apply IoU as an objective metric ( $M_{seg}$ ) in (11) throughout the WHS training for all networks, as mentioned above.

$$M_{seg}(y, \hat{y}) = \frac{\sum_{i=1}^N y_i \times \hat{y}_i}{\sum_{i=1}^N y_i + \sum_{i=1}^N \hat{y}_i - \sum_{i=1}^N y_i \times \hat{y}_i}, \quad (11)$$

where the parameters, such as  $y$ ,  $\hat{y}$ , and  $N$ , respectively indicate the actual label, prognosticated label, and the total voxel numbers. The multiplicative term of  $y$  and  $\hat{y}$  in the above equation is the estimation of identity (intersection) connecting the actual and predicted VOIs. The binary cross-entropy is utilized as the WHS's cost functions by the networks, such as EDNet, VGG-UNet, and VGG-FCN8s, optimized employing the Adam optimizer [61] with initial Learning Rate ( $LR$ ) and exponential decay rates ( $\beta_1$ ,  $\beta_2$ ) as  $LR = 0.001$ ,  $\beta_1 = 0.9$ , and  $\beta_2 = 0.999$ , respectively, without AMSGrad variant. The  $LR$  is decreased subsequent 5 epochs by 10.0% if validation loss obstructs progressing. The initial epochs are set as 200, which is suspend the network training utilizing a callback function after the validation loss becomes stagnated.

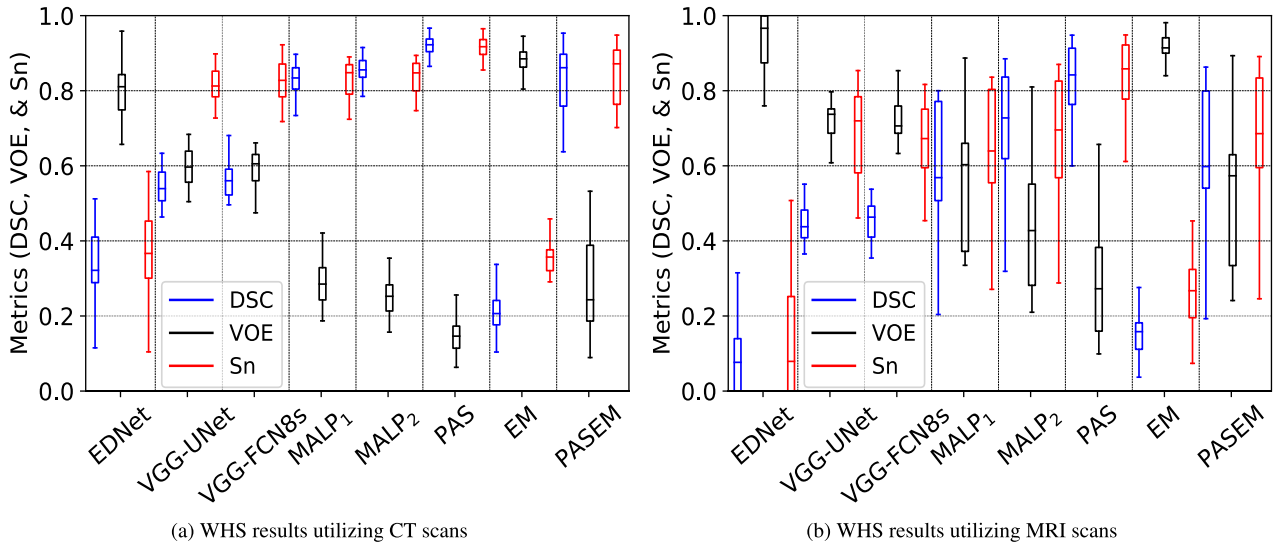
#### D. EVALUATION

The extensive experimentations are conducted on a computer operating on *Windows-10* system with the following hardware configuration: Intel® Core™ i7-7700 HQ CPU @ 2.80 GHz processor with Install memory (RAM): 32.0 GB and GeForce GTX 1080 GPU with 8 GB GDDR5 memory. The proposed pipeline and CNN-based models are designed utilizing the MATLAB R2020a and Python programming language with various Python and Keras APIs.

We utilize mean Volume Overlapping Error (mVOE), mean Sensitivity (mSn), and mean Dice Similarity Coeffi-

**TABLE 3.** The quantitative WHS results from our different methods utilizing both the CT and MRI scans. Best and second-best metrics are underlined and double-underlined, respectively. The smaller value of mVOE and the larger values of mSn and mDSC are the pointers of better WHS results.

Different methods	CT			MRI		
	mVOE	mSn	mDSC	mVOE	mSn	mDSC
EDNet	0.800 ± 0.080	0.384 ± 0.135	0.354 ± 0.126	0.937 ± 0.070	0.139 ± 0.154	0.095 ± 0.096
VGG-UNet	0.596 ± 0.054	0.796 ± 0.086	0.543 ± 0.048	0.720 ± 0.049	0.693 ± 0.120	0.444 ± 0.056
VGG-FCN8s	0.590 ± 0.053	0.828 ± 0.062	0.564 ± 0.052	0.719 ± 0.051	0.654 ± 0.128	0.450 ± 0.068
MALP	Median	0.291 ± 0.063	0.828 ± 0.051	0.556 ± 0.169	0.648 ± 0.163	0.595 ± 0.169
	MVF	<u>0.251 ± 0.051</u>	<u>0.835 ± 0.046</u>	<u>0.855 ± 0.033</u>	<u>0.445 ± 0.180</u>	<u>0.671 ± 0.169</u>
PAS	0.145 ± 0.046	0.919 ± 0.026	0.921 ± 0.027	0.291 ± 0.162	0.825 ± 0.126	0.819 ± 0.122
EM	0.878 ± 0.040	0.350 ± 0.063	0.216 ± 0.063	0.917 ± 0.038	0.260 ± 0.100	0.152 ± 0.064
PAS+EM	0.331 ± 0.144	0.804 ± 0.097	0.793 ± 0.107	0.527 ± 0.197	0.674 ± 0.187	0.619 ± 0.190

**FIGURE 7.** The Box and Whisker visualization of DSC (blue), VOE (black), and Sn (red) for the WHS utilizing (a) CT and (b) MRI scans employing eight different methods. The smaller VOE value and larger Sn and DSC values are the pointers of better WHS results. The MALP<sub>1</sub> and MALP<sub>2</sub> indicate a MALP method with the median and MVF schemes, respectively.

cient (mDSC) to quantify the segmentation accuracy, which are defined as follows:

$$mVOE = \frac{1}{S} \sum_{j=1}^S \left( 1 - \frac{TP_j}{TP_j + FN_j + FP_j} \right),$$

$$mSn = \frac{1}{S} \sum_{j=1}^S \frac{TP_j}{TP_j + FN_j},$$

$$mDSC = \frac{1}{S} \sum_{j=1}^S \frac{2 \cdot TP_j}{2 \cdot TP_j + FN_j + FP_j},$$

where  $S$ ,  $TP$ ,  $FN$ , and  $FP$  respectively denote the number of slices in an unseen test image, true-positive region (VOI as a VOI), false-negative area (VOI as a background), and false-positive area (background as a VOI). mSn determines the percentage of correctly segmented VOI of the true VOI, whereas mVOE and mDSC estimate the difference and similarity between true and segmented VOIs.

### III. RESULTS AND DISCUSSION

This section is dedicated to the presentation of extensive experimental results. In subsection III-A, we present

and interpret the results of eight different approaches (described in subsections II-B & II-C) utilizing the both CT and MRI scans. We have grouped those eight distinct methods into three categories: CNN-, MALP-, and probabilistic atlas-base methods, where each group has several methods. Firstly, we perform ablation studies among different methods in each group on the same dataset in subsections III-A1, III-A2, and III-A3, respectively. Finally, in subsection III-B, we analyze the best methods from all the groups for both the MRI and CT scans and compare seventeen separate published schemes with our best performing method on the same dataset.

#### A. EXPERIMENTAL RESULTS

Table 3 demonstrates the obtained results from different methods for the WHS on both the CT and MRI scans explicating various quantitative metrics (see in subsection II-D). Fig. 7 exhibits the Box and Whisker visualization of all the methods showing the spreads and centers of the DSC, VOE, and Sn metrics for all the 20 test images, either CT or MRI scans. Fig. 8 (a) and Fig. 8 (b) display the qualitative

segmentation results for both the CT (top) and MRI (bottom), respectively, from all the methods.

### 1) RESULTS FOR CNN-BASED METHODS

The experimental results in Table 3 quantitatively demonstrate that the EDNet is the worst-performing method than the other two CNN-based methods, such as VGG-UNet and VGG-FCN8s, by the margins of 18.9% and 21.0% for CT scans and 34.9% and 35.5% for MRI scans, respectively, concerning the mDSC. Those two networks also defeat the EDNet for the other two metrics (mVOE and mSn) with significant margins (see in Table 3). Although those three networks have an equal number of convolutional and pooling layers in the encoder and decoder, they constructionally vary in skip connection employment (see details in subsection II-C). The results on both the imaging modalities reveal that the appliance of skip connection has outputted better-segmented substructures, as the local information from the shallower layers is utilized to reconstruct output masks through the skip connection (see the results in Table 3). Again, it is seen from Fig. 7 that the DSC, VOE, and Sn from the EDNet are sparse for both the CT and MRI scans, with significantly fewer median metrics, which demonstrates that EDNet produces scattered results for each of the testing cases. On the other hand, the upper- and lower-whisker for all three metrics are closer for VGG-FCN8s and VGG-UNet than the EDNet (see in Fig. 7), which shows better-robustness of them comparing the EDNet.

Furthermore, the qualitative results in Fig. 8 depict that both the VGG-UNet and VGG-FCN8s provide better-segmented substructures of both the modalities (CT and MRI) than the EDNet. In several examples of the EDNet for MRI images, none of the target substructures are segmented, which provide VOE as 100.0% and Sn and DSC as 0.0%. Hence, the upper-whisker and upper-quartile of VOE are identical, which is 1.0 (see in Fig. 7 (b)). Similarly, the lower-whisker and lower-quartile of Sn and DSC are also equal, which are 0.0 (see in Fig. 7 (b)). The encoders in all three networks have 13-convolutional and 5-pooling layers, where the in-depth features obtained from them have lost spatial location information due to pooling in the encoders. Hence, the output masks from EDNet have less local information, which is solved in the VGG-UNet and VGG-FCN8s due to the concatenation of local information through the skip connection, providing an alternative path for the gradient with backpropagation. It is experimentally validated in our investigations that the additional skipping paths are beneficial for improving the segmentation results of different heart substructures.

Again, it is also remarkable that VGG-FCN8s even exceeds the VGG-UNet (see the corresponding results in Table 3, Fig. 7, and Fig. 8) by the margins of 2.1% and 0.6% respectively for CT and MRI scans in terms of mDSC. The former network further outperforms the latter network concerning the other two metrics, such as mVOE and mSn, with significant margins for the CT scans. The concatenation of low-level

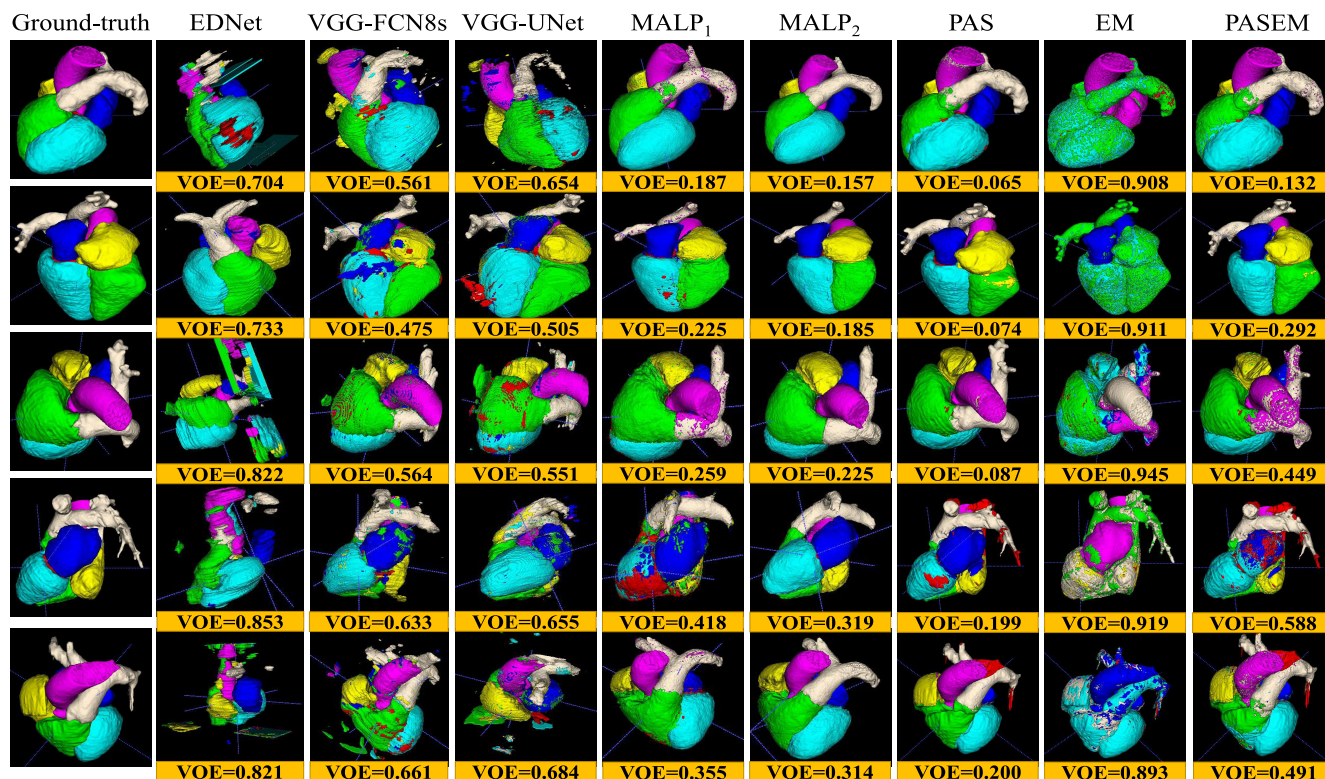
features from the encoder's antecedent layers with the equivalent decoder scale in VGG-UNet is the possible reason for failing VGG-UNet than VGG-FCN8s [16], [35]. Unessentially, an aggregation of the corresponding scale feature maps from the beginning layer of the encoder, is observed as a weakness of the UNet, as it imposes an undesirable merging procedure, forcing aggregation barely at the corresponding scale feature maps of the encoder and decoder, which was similarly experimentally validated in [35] and our previous article for other medical imaging modality [16]. On the other hand, in VGG-FCN8s, we fuse the feature maps from the encoder's different coarseness starting from the third block of the encoder, making it a winner of three implemented CNN-based WHS approaches.

### 2) RESULTS FOR MALP-BASED METHODS

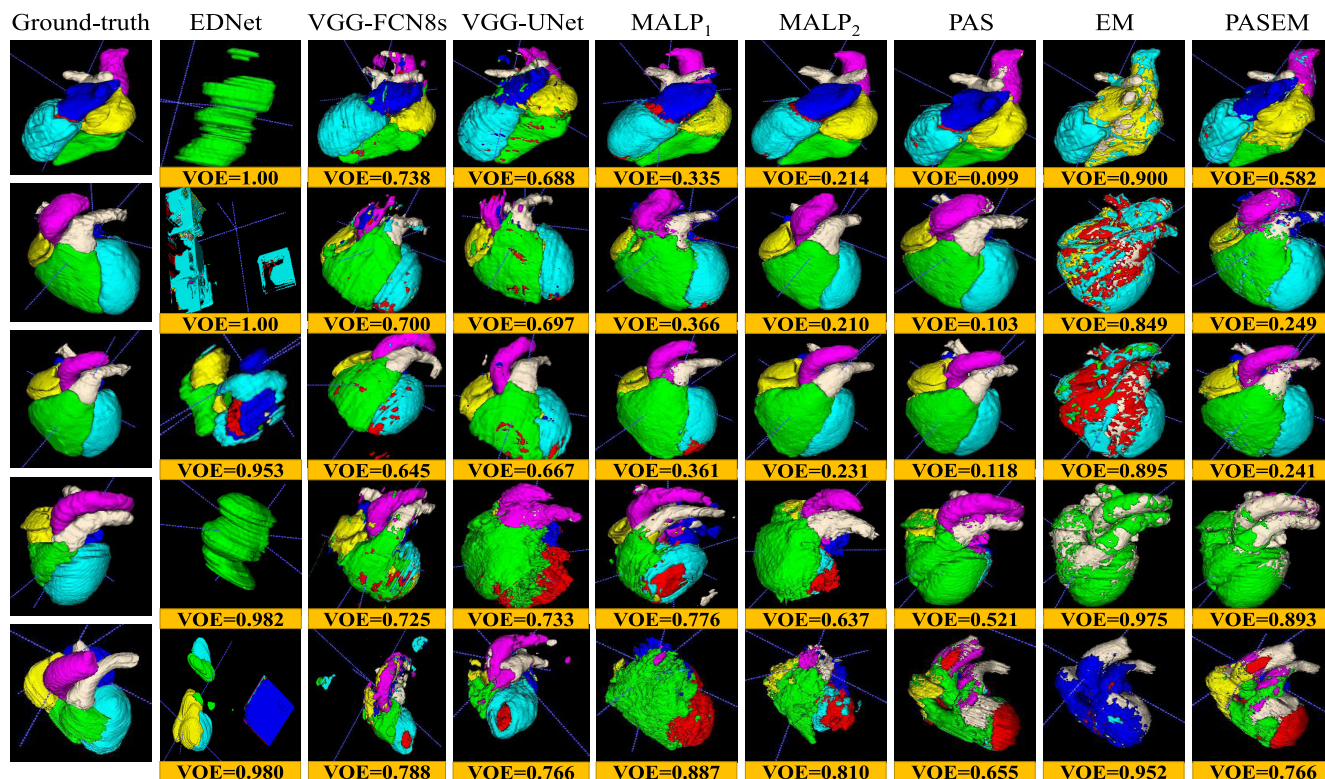
The label propagated segmentation results utilizing different deformed atlas images are quantitatively and qualitatively manifested in Table 3, Fig. 7, and Fig. 8. The median and MVF schemes of level propagation, as described in subsection II-B2, generate the heart segmentation results with the mVOEs of 29.1% and 25.1%, respectively, for CT scans and 55.6% and 44.5%, respectively, for MRI scans. Table 3 demonstrates that the MVF scheme outperforms the median method of MALP by the margins of 2.7% and 10.1% respectively for CT and MRI modalities for mDSC. The former MALP method also outperforms the latter MALP method with significant margins concerning the other two metrics (mVOE and mSn). The Box and Whisker visualization of all three metrics in Fig. 7 for both the methods demonstrate the superiority of the MVF scheme than the median strategy for both the imaging modalities. Fig. 8 (top) and Fig. 8 (bottom) exhibit the qualitative results respectively for CT and MRI scans for both the median (MALP<sub>1</sub>) and MVF (MALP<sub>2</sub>) schemes. Those results qualitatively confirm that the segmented substructures from MALP<sub>1</sub> suffer from the outliers (see in fifth and sixth columns of Fig. 8), where most of the organs are erroneously labeled with other organs. All the experimental results reveal that the MVF scheme has better dealt with the outliers as it counts the majority of the labels from the voting candidates, whereas the median method estimates the median values of the counter, which may not be matched by the majority voters.

### 3) RESULTS FOR PROBABILISTIC ATLAS-BASED METHODS

The WHS results of the probabilistic atlas are exhibited in Table 3, Fig. 7, and Fig. 8, where we employ our three methods, such as PAS, EM, and PAS+EM (see details in subsection II-B2). The PAS, EM, and PAS+EM schemes of probabilistic atlas provide the heart segmentation results with the mVOEs of 14.5%, 87.8%, and 33.1%, respectively, for CT scans and 29.1%, 91.7%, and 52.7%, respectively, for MRI scans. Table 3 exhibit that the PAS scheme exceeds the other two methods, such as EM and PAS+EM, by the margins of 70.5% and 12.8% for CT scans concerning the mDSC, respectively, whereas it also beats them by the margins of



(a) Qualitative WHS results in CT scans applying different methods



(b) Qualitative WHS results in MRI scans applying different methods

**FIGURE 8.** The qualitative WHS results from our different methods in CT (top) and MRI (bottom) scans, where the first three and last two rows of both the figures are for the best- and worst-performing examples for the best-performing method (PAS), respectively. The MALP<sub>1</sub> and MALP<sub>2</sub> indicate a MALP method with the median and MVF schemes, respectively.

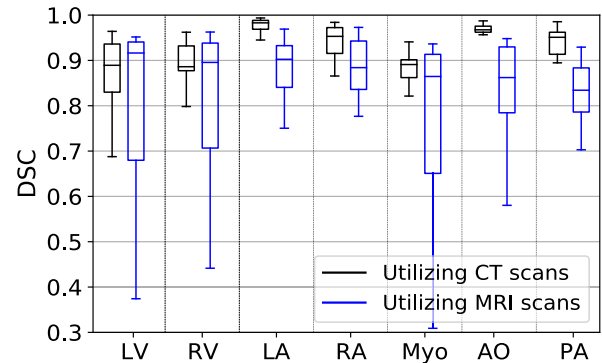
66.7 % and 20.0 % for MRI scans, respectively, in terms of the mDSC. Similarly, the EM and EM+PAS are also defeated by the proposed PAS method for mVOE and mSn for both the chest imaging modalities with considerable margins (see in Table 3). The spreads and centers of the DSC, VOE, and Sn metrics for all the 20 test images (either CT or MRI scans), as exhibited in Fig. 7, also reveal the supremacy of the proposed PAS method over the other two methods (EM and EM+PAS).

The proposed PAS technique estimates the voxel class and assigns a target voxel label from the posterior distribution of the Bayesian framework, where the posterior distribution is determined from the likelihood tissue model and prior anatomical knowledge by a MAP rule. Such an estimation of posterior distribution quantifies its expected probability value and the uncertainty associated with it, which probably makes the PAS algorithm a winner than the EM algorithm. As our EM algorithm is a single variate, it has less representation of the heart substructure features. The EM algorithm is prone to converging to local minima when dealing with a single variate dataset, which provides an overfitted model for heart segmentation.

However, incorporating the PAS algorithm with an EM algorithm extends the WHS outcomes by the margins of 57.7 % and 46.7 % in terms of mDSC respectively for CT and MRI modalities. Still, the PAS technique is a defeating method. Again, Fig. 8 (top) and Fig. 8 (bottom) show the qualitative WHS outcomes for the CT and MRI scans for the PAS, EM, and PAS+EM designs, respectively. Those results show that the segmented substructures from EM and PAS+EM yield the outliers substructures (see in eight and ninth columns of Fig. 8), where the EM results suffer severely by the outliers. The embodiment of PAS with the EM has significantly reduced the outlier effect from the WHS results of the EM technique, as qualitatively depicted in Fig. 8.

## B. RESULTS COMPARISON

Comparing all the results as mentioned earlier, the proposed PAS is the best performing method for the WHS for both the utilized dataset modalities, such as CT and MRI scans, which are manifested and visualized in Table 3, Fig. 7 and Fig. 8. The proposed and designed methods are classified into three categories: CNN-based, MALP-based, and probabilistic atlas-based, where the former approach works on the 2D slice of the 3D CT and MRI modalities, while the former two methods work on the whole 3D CT and MRI volumes. The experimentation reveals that the 3D image-based approach conquers the 2D image-based techniques. Our experimental WHS results also demonstrate that the 3D atlas-based approaches beat CNN-based 2D methods. This result's probable reasons are that we take only axial slices from the 3D CT and MRI scans to train and evaluate the network. Although the 2D CNNs have much lighter computation and higher inference speed, it neglects the information between adjacent slices, which hinders the improvement of segmentation accuracy. Hence, the 3D CNNs can be a powerful model for



**FIGURE 9.** The Box and Whisker visualization of DSC each substructure segmentation of WHS utilizing CT (black) and MRI (blue) scans. Our best performing segmentation method (PAS) for CT and MRI scans is employed to obtain this visualization.

learning representations for volumetric data and perceiving the volumetric spatial information, which will be analyzed for the same WHS task in the future.

Fig. 9 reflects the Box and Whisker presentation of the mDSC of our proposed PAS algorithm on both the CT and MRI modalities for seven different substructures of the heart (see subsection I-A). The CT scans produce the best WHS results than the MRI scans for all the substructures in terms of mean and median values of the mDSC. The substructure's segmentation results utilizing the MRI images also have a higher interquartile range than the CT images (see in Fig. 9), which indicates less robustness of the MRI images for the WHS. WHS's poor results utilizing the MRI images over CT images are also reported in previously published articles [1], [5], [30], [38]. The segmentation results, especially for LA, RA, Myo, AO, and PA, utilizing the CT images, are praiseworthy as they have a high median value and significantly less interquartile range than MRI images. Although MRI images' performance is less than CT, merging them can further improve the WHS results with the multi-modal input data, as it will have better feature presentations of different heart substructures [38]. Therefore, the future working direction can concentrate on the multi-modal utilization of the MM-WHS-2017 dataset applying this article's proposed techniques.

Table 4 shows the WHS results from our best performing method and other state-of-the-art methods on the same dataset utilizing both the CT and MRI scans. The comparative results, as in Table 4, demonstrate that the proposed PAS method with the CT scans outperforms the recent state-of-the-art approach in [38] by the margins of 1.3 % and 0.7 % concerning the mVOE and mDSC, respectively. The authors in the second-best method (MMLNet) [38] transferred the MRI image information from the source domain to the target CT domain through the adversarial training without considering their spatial alignment, which could be the viable reasons for realizing the WHS multi-modality approach. The other two well-performing methods (see in Table 4), such as Multi-label FCN [30] and Multi-depth fusion of 3D UNet

**TABLE 4. Comparative results of our best performing method for WHS and state-of-the-art techniques for the same task on the MRI, CT, and MRI+CT scans and same dataset. Best and second-best metrics are underlined and double-underlined, respectively. The smaller value of mVOE and larger values of mSn and mDSC are the pointers of better WHS results.**

Different methods	Year	Modalities	Metrics		
			mVOE	mSn	mDSC
Multi-atlas registration-based approach [62]	2017	MRI	0.246	—	0.860
Multi-planar CNN [5]	2017	MRI	0.259	0.831	0.851
Multi-label FCN [30]	2017	MRI	0.230	—	0.870
Averaging of 10 different algorithm [1]	2019	MRI	0.299	—	0.824
MvMM-RegNet: multivariate mixture model with registration framework [63]	2020	MRI	0.229	—	0.871
MMTLNet: multi-modality transfer learning with adversarial training [38]	2020	MRI	0.198	—	0.890
Multi-planar CNN [5]	2017	CT	0.187	<u>0.866</u>	0.897
Multi-label FCN [30]	2017	CT	0.168	—	0.908
CFUN: faster R-CNN + 3D UNet [19]	2018	CT	0.247	—	0.859
HFANet: deep heterogeneous feature aggregation network [28]	2019	CT	0.167	—	0.909
Multi-depth fusion of 3D UNet combining local and global features [34]	2019	CT	0.170	—	0.907
Averaging of 10 different algorithm [1]	2019	CT	0.227	—	0.872
MMTLNet: multi-modality transfer learning with adversarial training [38]	2020	CT	<u>0.158</u>	—	<u>0.914</u>
3D UNet incorporating of the principal component analysis for augmentation [64]	2020	CT	0.211	—	0.882
Multi-scale patch and multi-modality atlas-based WHS [15]	2016	CT+MRI	0.183	—	0.899
Two-stage UNet with adaptive threshold window [4]	2019	CT+MRI	0.343	—	0.793
Dual-Teacher: integrating intra-domain and inter-domain teachers [26]	2020	CT+MRI	0.245	—	0.860
<b>Our proposed probabilistic atlas-based WHS</b>	2020	MRI	0.291	0.825	0.819
<b>Our proposed probabilistic atlas-based WHS</b>	2020	CT	<u>0.145</u>	<u>0.919</u>	<u>0.921</u>

combining local and global features [34], also carry possible drawbacks for being defeated by our proposed approach. The former strategy applied two stages, wherein the first stage, the VOIs are selected for the essential second stage. The errors in VOI selection could lead to erroneous heart segmentation results. The latter approach employed the fusion from the different depths of the CNN network, where they fused the input block to the along with other depths to generate the final WHS VOIs. The concatenation of the input block with the same scale outermost decoder's block probably hamper the precise output, as it is also proven in our VGG-UNet and VGG-FCN8s experiments (see in Table 3), and the articles in [16], [35].

#### IV. CONCLUSION AND FUTURE WORK

Accurate in the whole heart's segmentation is crucial in developing clinical applications, although it is very challenging due to diverse artifacts, image variability, and patient-specific properties. This article has introduced and explored a robust and accurate pipeline for the WHS utilizing two different heart imaging modalities, such as CT and MRI scans. It is experimentally validated that incorporating the prior anatomical knowledge represented by probabilistic atlas into the Bayes inference to delineate seven different heart substructures has better-segmented results while utilizing the CT scans and employing our multi-resolution non-rigid registration pipeline. As the 2D CNNs fail to provide satisfactory WHS results, future research will focus our research direction on training and evaluating networks in multiple directions (coronal, sagittal, and axial) to combine all the directional WHS results. We will also design an end-to-end 3D segmentation network for comprehensive ablation studies. The recommended pipeline will be applied to other domains for volumetric medical image segmentation to verify its versatility and generability.

#### CONFLICT OF INTEREST

The authors have not any conflicts to disclose this research.

#### REFERENCES

- [1] X. Zhuang, "Evaluation of algorithms for multi-modality whole heart segmentation: An open-access grand challenge," *Med. Image Anal.*, vol. 58, Dec. 2019, Art. no. 101537.
- [2] S. Mendis, *Global Atlas on Cardiovascular Disease Prevention and Control*. Geneva, Switzerland: World Health Organization, 2011.
- [3] D. Kang, "Heart chambers and whole heart segmentation techniques: Review," *J. Electron. Imag.*, vol. 21, no. 1, Apr. 2012, Art. no. 010901.
- [4] T. Liu, Y. Tian, S. Zhao, X. Huang, and Q. Wang, "Automatic whole heart segmentation using a two-stage U-net framework and an adaptive threshold window," *IEEE Access*, vol. 7, pp. 83628–83636, 2019.
- [5] A. Mortazi, J. Burt, and U. Bagci, "Multi-planar deep segmentation networks for cardiac substructures from MRI and CT," in *Proc. Int. Workshop Stat. Atlases Comput. Models Heart*. Berlin, Germany: Springer, 2017, pp. 199–206.
- [6] Q. Dou, L. Yu, H. Chen, Y. Jin, X. Yang, J. Qin, and P.-A. Heng, "3D deeply supervised network for automated segmentation of volumetric medical images," *Med. Image Anal.*, vol. 41, pp. 40–54, Oct. 2017.
- [7] T. Heimann, "Comparison and evaluation of methods for liver segmentation from CT datasets," *IEEE Trans. Med. Imag.*, vol. 28, no. 8, pp. 1251–1265, Aug. 2009.
- [8] A. Radtke, S. Nadalin, G. C. Sotiropoulos, E. P. Molmenti, T. Schroeder, C. Valentin-Gamazo, H. Lang, M. Bockhorn, H. O. Peitgen, C. E. Broelsch, and M. Malago, "Computer-assisted operative planning in adult living donor liver transplantation: A new way to resolve the dilemma of the middle hepatic vein," *World J. Surg.*, vol. 31, no. 1, p. 175, 2007.
- [9] H.-P. Meinzer, M. Thorn, and C. E. Cárdenas, "Computerized planning of liver surgery—An overview," *Comput. Graph.*, vol. 26, no. 4, pp. 569–576, 2002.
- [10] J. Peters, O. Ecabert, C. Meyer, H. Schramm, R. Kneser, A. Groth, and J. Weese, "Automatic whole heart segmentation in static magnetic resonance image volumes," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Berlin, Germany: Springer, 2007, pp. 402–410.
- [11] D. F. Pace, A. V. Dalca, T. Geva, A. J. Powell, M. H. Moghari, and P. Golland, "Interactive whole-heart segmentation in congenital heart disease," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Springer, 2015, pp. 80–88.

- [12] A. Atehortúa, M. A. Zuluaga, S. Ourselin, D. Giraldo, and E. Romero, "Automatic segmentation of 4D cardiac mr images for extraction of ventricular chambers using a spatio-temporal approach," *Proc. SPIE Med. Imag. Image Process.*, vol. 9784, May 2016, Art. no. 978435.
- [13] F. Zhao, H. Hu, Y. Chen, J. Liang, X. He, and Y. Hou, "Accurate segmentation of heart vol. in, cta with landmark-based registration and fully convolutional network," *IEEE Access*, vol. 7, pp. 57881–57893, 2019.
- [14] V. H. C. de Albuquerque, D. de A. Rodrigues, R. F. Ivo, S. A. Peixoto, T. Han, W. Wu, and P. P. Rebouças Filho, "Fast fully automatic heart fat segmentation in computed tomography datasets," *Comput. Med. Imag. Graph.*, vol. 80, Mar. 2020, Art. no. 101674.
- [15] X. Zhuang and J. Shen, "Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI," *Med. Image Anal.*, vol. 31, pp. 77–87, Jul. 2016.
- [16] M. K. Hasan, M. A. Alam, M. T. E. Elahi, S. Roy, and R. Martí, "DRNet: Segmentation and localization of optic disc and fovea from diabetic retinopathy image," *Artif. Intell. Med.*, vol. 111, Jan. 2021, Art. no. 102001.
- [17] M. K. Hasan, L. Dahal, P. N. Samarakoon, F. I. Tushar, and R. Martí, "DSNet: Automatic dermoscopic skin lesion segmentation," *Comput. Biol. Med.*, vol. 120, May 2020, Art. no. 103738.
- [18] D. M. Vigneault, W. Xie, C. Y. Ho, D. A. Bluemke, and J. A. Noble, "Ω-net (omega-net): Fully automatic, multi-view cardiac MR detection, orientation, and segmentation with deep neural networks," *Med. Image Anal.*, vol. 48, pp. 95–106, Aug. 2018.
- [19] Z. Xu, Z. Wu, and J. Feng, "CFUN: Combining faster R-CNN and U-net network for efficient whole heart segmentation," 2018, *arXiv:1812.04914*. [Online]. Available: <http://arxiv.org/abs/1812.04914>
- [20] X. Ding, Y. Peng, C. Shen, and T. Zeng, "Cab U-net: An end-to-end category attention boosting algorithm for segmentation," *Comput. Med. Imag. Graph.*, vol. 84, May 2020, Art. no. 101764.
- [21] T. Han, R. F. Ivo, D. D. A. Rodrigues, S. A. Peixoto, V. H. C. de Albuquerque, and P. P. Rebouças Filho, "Cascaded volumetric fully convolutional networks for whole-heart and great vessel 3D segmentation," *Future Gener. Comput. Syst.*, vol. 108, pp. 198–209, Jul. 2020.
- [22] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum, "Dilated convolutional neural networks for cardiovascular mr segmentation in congenital heart disease," in *Reconstruction, Segmentation, Analysis Medical Images*. Berlin, Germany: Springer, 2016, pp. 95–102.
- [23] M. A. Zuluaga, B. Biffi, A. M. Taylor, S. Schievano, T. Vercauteren, and S. Ourselin, "Strengths and pitfalls of whole-heart atlas-based segmentation in congenital heart disease patients," in *Reconstruction, Segmentation, Analysis Medical Images*. Berlin, Germany: Springer, 2016, pp. 139–146.
- [24] C. Wang, Q. Wang, and O. Smedby, "Automatic heart and vessel segmentation using random forests and a local phase guided level set method," in *Reconstruction, Segmentation, Analysis Medical Images*. Berlin, Germany: Springer, 2016, pp. 159–164.
- [25] L. Yu, X. Yang, J. Qin, and P.-A. Heng, "3D FractalNet: Dense volumetric segmentation for cardiovascular MRI volumes," in *Reconstruction, Segmentation, Analysis Medical Images*. Berlin, Germany: Springer, 2016, pp. 103–110.
- [26] K. Li, S. Wang, L. Yu, and P.-A. Heng, "Dual-teacher: Integrating intra-domain and inter-domain teachers for annotation-efficient cardiac segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Berlin, Germany: Springer, 2020, pp. 418–427.
- [27] M. Habijan, H. Leventić, I. Galić, and D. Babin, "Whole heart segmentation from CT images using 3D U-net architecture," in *Proc. Int. Conf. Syst., Signals Image Process. (IWSSIP)*, Jun. 2019, pp. 121–126.
- [28] H. Zheng, L. Yang, J. Han, Y. Zhang, P. Liang, Z. Zhao, C. Wang, and D. Z. Chen, "HFA-net: 3D cardiovascular image segmentation with asymmetrical pooling and content-aware fusion," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Berlin, Germany: Springer, 2019, pp. 759–767.
- [29] X. Dong, Y. Lei, T. Wang, M. Thomas, L. Tang, W. J. Curran, T. Liu, and X. Yang, "Automatic multiorgan segmentation in thorax CT images using U-net-gan," *Med. Phys.*, vol. 46, no. 5, pp. 2157–2168, 2019.
- [30] C. Payer, D. Štern, H. Bischof, and M. Urschler, "Multi-label whole heart segmentation using CNNs and anatomical label configurations," in *Proc. Int. Workshop Stat. Atlases Comput. Models Heart*. Berlin, Germany: Springer, 2017, pp. 190–198.
- [31] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Berlin, Germany: Springer, 2015, pp. 234–241.
- [32] C. Wang and O. Smedby, "Automatic whole heart segmentation using deep learning and shape context," in *Proc. Int. Workshop Stat. Atlases Comput. Models Heart*. Berlin, Germany: Springer, 2017, pp. 242–249.
- [33] Q. Tong, M. Ning, W. Si, X. Liao, and J. Qin, "3D deeply-supervised U-net based whole heart segmentation," in *Proc. Int. Workshop Stat. Atlases Comput. Models Heart*. Berlin, Germany: Springer, 2017, pp. 224–232.
- [34] C. Ye, W. Wang, S. Zhang, and K. Wang, "Multi-depth fusion network for whole-heart CT image segmentation," *IEEE Access*, vol. 7, pp. 23421–23429, 2019.
- [35] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 1856–1867, Jun. 2020.
- [36] Y. Mo, F. Liu, D. McIlwraith, G. Yang, J. Zhang, T. He, and Y. Guo, "The deep poincaré map: A novel approach for left ventricle segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Berlin, Germany: Springer, 2018, pp. 561–568.
- [37] A. Gubern-Mérida, M. Kallenberg, R. Martí, and N. Karssemeijer, "Multi-class probabilistic atlas-based segmentation method in breast MRI," in *Proc. Iberian Conf. Pattern Recognit. Image Anal.* Berlin, Germany: Springer, 2011, pp. 660–667.
- [38] X. Liao, Y. Qian, Y. Chen, X. Xiong, Q. Wang, and P.-A. Heng, "MMTL-Net: Multi-modality transfer learning network with adversarial training for 3D whole heart segmentation," *Comput. Med. Imag. Graph.*, vol. 85, Oct. 2020, Art. no. 101785.
- [39] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: Application to breast MR images," *IEEE Trans. Med. Imag.*, vol. 18, no. 8, pp. 712–721, Oct. 1999.
- [40] C. R. Maurer, D. L. G. Hill, A. J. Martin, H. Liu, M. McCue, D. Rueckert, D. Lloret, W. A. Hall, R. E. Maxwell, D. J. Hawkes, and C. L. Truwit, "Investigation of intraoperative brain deformation using a 1.5-T interventional MR system: Preliminary results," *IEEE Trans. Med. Imag.*, vol. 17, no. 5, pp. 817–825, Jun. 1998.
- [41] M. Lorenzo-Valdés, G. I. Sanchez-Ortiz, R. Mohiaddin, and D. Rueckert, "Atlas-based segmentation and tracking of 3D cardiac mr images using non-rigid registration," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Berlin, Germany: Springer, 2002, pp. 642–650.
- [42] P. H. C. Eilers and B. D. Marx, "Flexible smoothing with B-splines and penalties," *Stat. Sci.*, vol. 11, no. 2, pp. 89–102, May 1996.
- [43] S. Klein, J. P. W. Pluim, M. Staring, and M. A. Viergever, "Adaptive stochastic gradient descent optimisation for image registration," *Int. J. Comput. Vis.*, vol. 81, p. 227, Mar. 2009.
- [44] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank, "PET-CT image registration in the chest using free-form deformations," *IEEE Trans. Med. Imag.*, vol. 22, no. 1, pp. 120–128, Jan. 2003.
- [45] I. Išgum, M. Staring, A. Rutten, M. Prokop, M. A. Viergever, and B. Van Ginneken, "Multi-atlas-based segmentation with local decision fusion—Application to cardiac and aortic segmentation in CT scans," *IEEE Trans. Med. Imag.*, vol. 28, no. 7, pp. 1000–1010, Jan. 2009.
- [46] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. Pluim, "Elastix: A toolbox for intensity-based medical image registration," *IEEE Trans. Med. Imag.*, vol. 29, no. 1, pp. 196–205, Jan. 2010.
- [47] A. Klein and J. Hirsch, "Mindboggle: A scatterbrained approach to automate brain labeling," *NeuroImage*, vol. 24, no. 2, pp. 261–280, Jan. 2005.
- [48] C. Ciofolo and C. Barillot, "Atlas-based segmentation of 3D cerebral structures with competitive level sets and fuzzy control," *Med. Image Anal.*, vol. 13, no. 3, pp. 456–470, Jun. 2009.
- [49] Y. Wu, S. K. Warfield, I. L. Tan, W. M. Wells, D. S. Meier, R. A. van Schijndel, F. Barkhof, and C. R. G. Guttmann, "Automated segmentation of multiple sclerosis lesion subtypes with multichannel MRI," *NeuroImage*, vol. 32, no. 3, pp. 1205–1215, Sep. 2006.

- [50] M. Cabezas, A. Oliver, X. Lladó, J. Freixenet, and M. B. Cuadra, "A review of atlas-based segmentation for magnetic resonance brain images," *Comput. Methods Programs Biomed.*, vol. 104, no. 3, pp. e158–e177, 2011.
- [51] I. S. Amiri, O. A. Akanbi, and E. Fazeldehkordi, *A Machine-Learning Approach To Phishing Detection Defense*. Rockland, MA, USA: Syngress, 2014.
- [52] H. Park, P. H. Bland, and C. R. Meyer, "Construction of an abdominal probabilistic atlas and its application in segmentation," *IEEE Trans. Med. Imag.*, vol. 22, no. 4, pp. 483–492, Apr. 2003.
- [53] N. Tajbakhsh, L. Jeyaseelan, Q. Li, J. N. Chiang, Z. Wu, and X. Ding, "Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation," *Med. Image Anal.*, vol. 63, Jul. 2020, Art. no. 101693.
- [54] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.
- [55] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, P. Martinez-Gonzalez, and J. Garcia-Rodriguez, "A survey on deep learning techniques for image and video semantic segmentation," *Appl. Soft Comput.*, vol. 70, pp. 41–65, Sep. 2018.
- [56] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: <http://arxiv.org/abs/1409.1556>
- [57] 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. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
- [58] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: <http://arxiv.org/abs/1502.03167>
- [59] A. Odena, V. Dumoulin, and C. Olah, "Deconvolution and checkerboard artifacts," *Distill*, vol. 1, no. 10, p. e3, 2016.
- [60] A. Rasmus, M. Berglund, M. Honkala, H. Valpola, and T. Raiko, "Semi-supervised learning with ladder networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 3546–3554.
- [61] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: <http://arxiv.org/abs/1412.6980>
- [62] M. P. Heinrich and J. Oster, "MRI whole heart segmentation using discrete nonlinear registration and fast non-local fusion," in *Proc. Int. Workshop Stat. Atlases Comput. Models Heart*. Berlin, Germany: Springer, 2017, pp. 233–241.
- [63] X. Luo and X. Zhuang, "MvMM-RegNet: A new image registration framework based on multivariate mixture model and neural network estimation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Berlin, Germany: Springer, 2020, pp. 149–159.
- [64] M. Habijan, H. Leventiá, I. Galíá, and D. Babin, "Neural network based whole heart segmentation from 3D CT images," *Int. J. Electr. Comput. Eng. Syst.*, vol. 11, no. 1, pp. 25–31, Apr. 2020.



**MD. KAMRUL HASAN** was born in Tangail, Bangladesh, in 1992. He received the B.Sc. and M.Sc. degrees in electrical and electronic engineering (EEE) from the Khulna University of Engineering and Technology (KUET), in 2014 and 2017, respectively, and the M.Sc. degree in medical imaging and application (MAIA) from the University of Burgundy, France, the University of Cassino and Southern Lazio, Italy, and the University of Girona, Spain, as an Erasmus Scholar, in 2019. He is currently working as an Assistant Professor with the Electrical and Electronic Engineering Department, KUET. He analyzed different medical image modalities and machine learning during the MAIA study to build a generic computer-aided diagnosis system. He is also a supervisor of several undergraduate students on the classification, segmentation, and registration of medical images with different modalities. He has already published many research articles on medical image and signal processing in different international journals and conferences. His research interests include medical image and data analysis, machine learning, deep convolutional neural networks, medical image reconstruction, and surgical robotics in minimally invasive surgery.



**SHIDHARTH ROY** (Student Member, IEEE) received the B.Sc. degree in electrical and electronic engineering from the Khulna University of Engineering and Technology (KUET), in 2020. His undergraduate thesis was on EEG Based Brain-Computer Interaction with Different Stimuli Using Machine Learning. He is currently working on bio-signal processing, medical imaging, 3D reconstruction of medical images, renewables, and computer-assisted interventions. His previous works were presented at TENSYP 2020, ICAEE 2019, ICIECE 2019, EICT 2019, and *Artificial Intelligence in Medicine* (AIMM, Elsevier). His research interests include artificial intelligence, brain-computer interface, and deep learning. He awarded seven national awards so far for his work.



**TARUN KANTI GHOSH** was born in Jashore, Khulna, Bangladesh, in 1983. He received the B.S. and M.S. degrees in biotechnology and genetic engineering from the Khulna University of Khulna, Bangladesh, in 2007 and 2011, respectively, and the M.S. degree in biomedical engineering from the Khulna University of Engineering and Technology (KUET), Khulna, in 2016, where he is currently pursuing the Ph.D. degree. From 2006 to 2008, he was a Research Fellow with the Animal Cell Culture and Molecular Biology Laboratory. He is the author of three international journal articles and conference papers. His research interests include biomedical signal and image processing, bioelectricity, bioinformatics, biomedical systems modeling and simulation, and computational intelligence in biomedical engineering.



**MD. ASHRAFUL ALAM** is currently pursuing the degree in electrical and electronic engineering (EEE) with the Khulna University of Engineering and Technology (KUET). He is also working on medical imaging. His previous works were presented at IEEE Access and *Artificial Intelligence in Medicine* (AIMM, Elsevier). His research interests include medical image and data processing, computer vision, and deep learning. He awarded two national awards so far for his work in the idea development project.





**EKLAS HOSSAIN** (Senior Member, IEEE) received the B.S. degree in electrical and electronic engineering from the Khulna University of Engineering and Technology, Bangladesh, in 2006, the M.S. degree in mechatronics and robotics engineering from the International Islamic University of Malaysia, Malaysia, in 2010, and the Ph.D. degree from the College of Engineering and Applied Science, University of Wisconsin Milwaukee (UWM). He has been working in distributed power systems and renewable energy integration for the last ten years.

Since 2015, he has been an Assistant Professor with the Department of Electrical Engineering and Renewable Energy, Oregon Tech, involved with several research projects on renewable energy and grid-tied microgrid system. He has published several research articles and posters in this field. He is also working as an Associate Researcher with the Oregon Renewable Energy Center (OREC). He is a registered Professional Engineer (PE) in the state of Oregon, USA. He is also a Certified Energy Manager (CEM) and Renewable Energy Professional (REP). His research interests include modeling, analysis, design, and control of power electronic devices, energy storage systems, renewable energy sources, integration of distributed generation systems, microgrid and smart grid applications, robotics, and advanced control systems. He is a Senior Member of the Association of Energy Engineers (AEE). He is also serving as an Associate Editor for IEEE ACCESS. He is the winner of the Rising Faculty Scholar Award, in 2019, from the Oregon Institute of Technology for his outstanding contribution to teaching. With his dedicated research team, he is looking forward to exploring methods to make electric power systems more sustainable, cost-effective, and secure through extensive research and analysis on energy storage, microgrid system, and renewable energy sources.



**MOHIUDDIN AHMAD** (Member, IEEE) received the B.S. degree (Hons.) in electrical and electronic engineering (EEE) from the Chittagong University of Engineering and Technology (CUET), Bangladesh, in 1994, the M.S. degree in electronics and information science (major – biomedical engineering) from the Kyoto Institute of Technology of Japan, in 2001, and the Ph.D. degree in computer science and engineering (CSE) from Korea University, South Korea, in 2008. From

November 1994 to August 1995, he worked as a part-time Lecturer with the Department of EEE, CUET. From August 1995 to October 1998, he worked as a Lecturer with the Department of EEE, Khulna University of Engineering and Technology (KUET), Bangladesh. He joined the Department of EEE, KUET, as an Assistant Professor, in June 2001. He joined the Department of EEE, KUET, as an Associate Professor, in May 2009, where he is currently a Full Professor. From October 2009 to September 2012, he worked as the Head of the Department of Biomedical Engineering (BME). From September 2012 to August 2014, he worked as the Head of the Department of EEE. Since July 2014, he has been serving as the sub-project manager of the UGC, HEQEP, Sub-Project, CP#3472, titled on Postgraduate Research in BME. His research interests include biomedical signal and image processing for disease diagnosis, computer vision and pattern recognition, clinical engineering, and modern healthcare. He is a Life Fellow of IEB.

• • •