IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 14, 2021, accepted June 7, 2021, date of publication June 14, 2021, date of current version June 21, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3088946

Semantic Segmentation of Cerebellum in 2D Fetal Ultrasound Brain Images Using Convolutional Neural Networks

VISHAL SINGH[1], PRADEEBA SRIDAR[2], JINMAN KIM[3], (Member, IEEE), RALPH NANAN[2], N. POORNIMA[4], SHANMUGA PRIYA[4], G. SAMEERA REDDY[4], SATHYABAMA CHANDRASEKARAN[4], AND RAMARATHNAM KRISHNAKUMAR[1]

[1]Department of Engineering Design, IIT Madras, Chennai 600036, India
[2]Sydney Medical School Nepean, The University of Sydney, Sydney, NSW 2747, Australia
[3]School of Computer Science, The University of Sydney, Sydney, NSW 2006, Australia
[4]Athena Diagnostics Imaging Centre, Chennai 600018, India

Corresponding author: Pradeeba Sridar (pradeeba.sridar@sydney.edu.au)

This work was supported by the Scheme for Promotion of Academic and Research Collaboration (SPARC) research grants provided by the Government of India.

ABSTRACT Cerebellum measurements of routinely acquired ultrasound (US) images are commonly used to estimate gestational age and to assess structural abnormalities of the developing central nervous system. Investigating associations between the developing cerebellum and neurodevelopmental outcomes post partum requires standardized cerebellum measurements from large clinical datasets. Such investigations have the potential to identify structural changes that can be used as biomarkers to predict growth and neurodevelopmental outcomes. For this purpose, high throughput, accurate, and unbiased measurements are necessary to replace existing manual, semi-automatic, and automated approaches which are tedious and lack reproducibility and accuracy. In this study, we propose a new deep learning algorithm for automated segmentation of the fetal cerebellum from 2-dimensional (2D) US images. We propose ResU-Net-c a semantic segmentation model optimized for fetal cerebellum structure. We leverage U-Net as a base model with the integration of residual blocks (Res) and introduce dilation convolution in the last two layers to segment the cerebellum (c) from noisy US images. Our experiments used a 5-fold cross-validation with 588 images for training and 146 for testing. Our ResU-Net-c achieved a mean Dice Score Coefficient, Hausdorff Distance, Recall, and Precision of 87.00%, 28.15, 86.00%, and 90.00%, respectively. The superiority of the proposed method over the other U-Net based methods is statistically significant ($p < 0.001$). Our proposed method can be leveraged to enable high throughput image analysis in clinical research fetal US images and can be employed in the biometric assessment in fetal US images on a larger scale.

INDEX TERMS Convolutional neural networks, fetal cerebellum, ResU-Net, segmentation, ultrasound images.

I. INTRODUCTION

Ultrasound (US) imaging is a routinely used modality to monitor fetal growth and development. Measurement of fetal brain structures includes the cerebellum, cerebrum, midbrain, and thalamus on US images, and it forms a part of the fetal anomaly screening performed at 18-21 weeks gestation. Studies have found that alterations in cerebellum development are linked to neurodevelopmental impairments involving general motor function, mental development, and disorders such as autism [1]–[6]. The cerebellum is highly conserved in its developmental stages, clearly demarcated from surrounding brain structures and hence easy to evaluate on routine US images. This makes the cerebellum an important target structure to understand neurodevelopmental outcomes and identify perturbations in the antenatal period that affect its development. Current clinical practices for the measuring of the cerebellum from US images are based on manual or semi-automatic techniques. Manual measurements require free-hand annotation by an experienced clinician, whereas

The associate editor coordinating the review of this manuscript and approving it for publication was Kumaradevan Punithakumar.

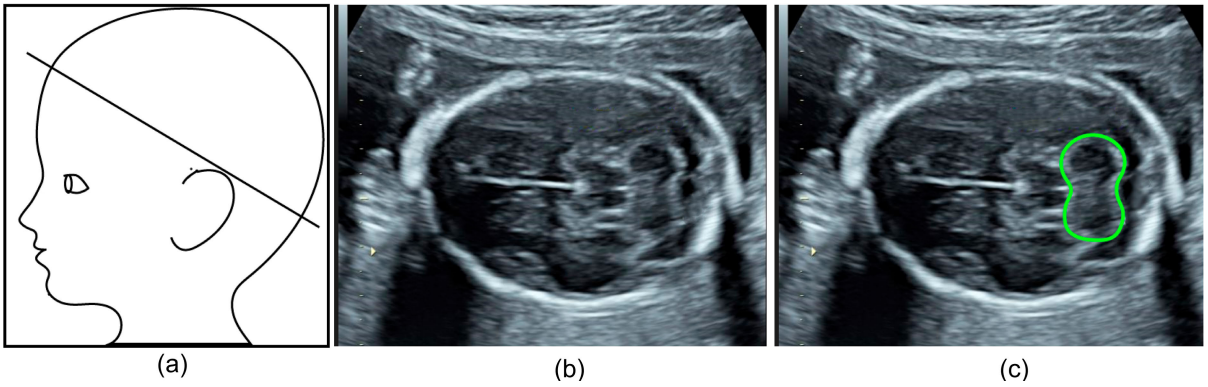

FIGURE 1. Representative of trans-cerebellar plane of an US image. (a) A schematic showing the cross-sectional (line) view acquired by an US scanner. (b) Trans-cerebellar plane. (c) Cerebellum depicted in green.

semi-automatic techniques involve user input to fix the 'end points' of the cerebellum, which are used by an automated algorithm to produce measurements. However, both of these approaches are time-consuming and require considerable clinical expertise, as these assessments involve subtle measurements of the width of the cerebellum from US images comprising of varying sizes based on the presentations of the fetus. In addition, US images inherently have signal drop out, motion artifacts and non-uniform contrast resolution.

Fig. 1 shows a sample representation of the trans-cerebellar plane of a US image. Semi-automatic approaches are often followed through manual corrections by an expert to rectify the external boundaries of the cerebellum. Manual measurement of the cerebellum depends on the sonographer's experience and is often subjected to inter- and intra-observer variability. The manual investigation of US images is also a time-consuming and tedious process, especially for larger sample sizes. Hence, it becomes important to develop automated US image analysis methods to overcome the need for user interaction and inter-observer variability.

There have been several attempts at automating the cerebellum measurements. For accurate cerebellum measurements, segmentation of the cerebellum region acts as a prior. In three-dimensional (3D) US volumes, Liu *et al.* [7], used weighted Hough transform and a constrained randomized Hough transform to detect the fetal brain midline and the skull. This anatomical information was used to identify the location of the cerebellum using a constrained probabilistic boosting tree. Yaqub *et al.* [8] used Random decision forest to segment the four fetal structures, including the cerebellum and suggested the requirement of shape constraints to segment regions that lack definite boundaries. Although the 3D US is better in depicting the fetal brain structures, in clinical practice, 2D US images are the dominant imaging modality due to their availability and standardization of the clinical protocol [9]. With 2D US images, although certain algorithms have imposed specific shape constraints to segment the cerebellum [10], they exhibit high computational load and do not provide satisfactory accuracy.

In recent years, segmentation algorithms based on convolutional neural network (CNN) have become state-of-the-art. In particular, a feed forward network was set up to build a U-shaped architecture and this has been successfully applied to medical image segmentation [11]. U-Net based CNN architecture was used to segment multiple brain structures including the cerebellum using weak labels in 3D US volumes [12]. Apart from 3D US image segmentation, Ramos *et al.* [13] applied YOLO (You Only Look Once) architecture to detect the cerebellum region with bounding boxes in 2D US images. However, this method is for localization of the cerebellum and evaluated on a smaller sample of images. Nevertheless, semantic segmentation is a preferred pre-requisite for automated fetal biometry; Semantic segmentation refers to a process of linking each pixel in an image to a class of cerebellum.

In this study, we present a semantic segmentation technique based on deep CNN for automatic segmentation of the cerebellum in US images. Our segmentation method, ResU-Net-c proposes the use of U-Net [11] architecture, and introduces residual blocks (Res) with dilated convolutions units in the last two layers. Both the Res and dilated convolutions modules are optimized to retain the spatial resolution of the US images. Thereby, the subtle structure of the fetal cerebellum (c) is segmented from noisy US images.

The contribution of our work lies in the following,

1) This is the first study to report on the semantic segmentation of the cerebellum structure in 2D fetal US images.
2) ResU-Net-c is an efficient semantic segmentation network, where it leverages the strength of both deep residual learning and U-Net architecture. The dilated convolutional layers increases the receptive field of the network without losing the spatial resolution. This module enables the network to learn the low and high-level image features to aid in the segmentation of subtle structures in the fetal brain, thereby overcoming the inherent limitation of the US image characteristics in the segmentation task.

3) We benchmarked our algorithm in comparison to the state-of-the-art on a large image dataset.

II. MATERIALS

The images were acquired using a GE Voluson E8 US machine from the Athena Diagnostics Imaging Centre, Chennai, India. All images were de-identified. The images were obtained using a standard-of-care clinical protocol that is a part of the routine clinical care; experienced qualified specialist radiologists followed the protocol established by the International Society of Ultrasound in Obstetrics & Gynaecology Education Committee (2007) for imaging fetal brain structures during the morphology scan at 18-20 weeks. We obtained a data set of 734 2D fetal US images of fetal trans-cerebellar plane.

Our experiments used a 5-fold cross-validation. Of the 734 images, each fold used 588 images for training and 146 for testing. The images were obtained in TIFF format. We cropped the images to 720 × 720 pixels, focusing on the cerebellum region, to fit the images to the segmentation network. The ground truth for the cerebellum region (segmentation masks) was prepared by a researcher under clinical supervision and was validated by clinicians.

III. METHODS

A. U-NET

U-Net is a deep CNN proposed by Ronneberger *et al.* [11] for semantic segmentation of biomedical images. It aims to label each pixel in the image with a corresponding label class. U-Net improves upon the fully CNN architecture [14] by expanding the decoder module's capacity. The U-Net architecture consists of a contracting path to capture context and a symmetric expanding path that enables precise localization. In the proposed method, U-Net serves as a backbone of the architecture with encoder and decoder paths.

B. RESIDUAL NETWORKS

Usage of deep layers in CNN evinces that they progressively learn more complex features. This can be beneficial to discriminately learn the complex image features of the cerebellum. However, the deeper networks have higher training and test error. He *et al.* [15], showed that overfitting could be caused due to the creation of complex functions with more layers. This may be the reason for the failure of deeper networks compared to shallow networks. The problem of overfitting can be suppressed with the use of an additional algorithm and regularization parameters. However, the failure of deeper networks is also attributed to the vanishing gradient problem due to the vast exploration of the feature space. This makes it prone to perturbations that may cause it to leave the manifold, and require additional labeled training data to recover, which is difficult to obtain in the medical imaging community.

The problem of training a very deep network has been alleviated with the use of residual neural networks. The Residual neural network makes use of skip connections to jump over some layers. ResNet models are implemented with double or triple layer skips. These layers usually contain rectified learning unit (ReLU) and batch normalization. The motivation for skipping the layers is to avoid vanishing gradients by reusing weights learned by the activation layer. Moreover, skipping simplifies the network by using fewer layers in the training stages and avoids the need for large training data. During training, weights adapt to the mute upstream layer, and amplify the previously skipped layer. As it learns the feature space, the network gradually restores the skipped layers. When all layers are extended, it stays closer to the manifold at the end of the training, resulting in faster learning. Skip connections have been shown to increase the performance in several image recognition tasks with deep networks. In our task, we have adapted skip connections to improve the performance of the task of segmentation.

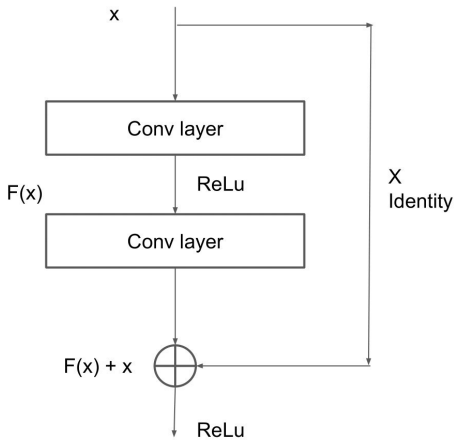

FIGURE 2. A basic residual block.

A basic residual block is shown in Fig. 2. The identity mapping does not have any parameters and is used only to add the output from the previous layer to the next layer. The dimensions of x and F(x) may not be the same. The identity mapping is multiplied by a linear projection W_s to expand the skip channel to match the dimensions.

$$y = F(x, \{W_i\}) + W_s x \tag{1}$$

The function $F(x, \{W_i\})$ represents the residual mapping to be learned. The W_s term can implement 'n' convolutions, we have introduced a convolutional layer (1 × 1) in ResU-Net-c.

C. ResU-NET-C ARCHITECTURE

The architecture of ResU-Net-c is presented in Fig. 3(a). It takes an input of size 720 × 720 pixels and generates a corresponding label map of the same size. The network is divided into two parts: an encoder (contracting path) in the left half and a decoder (expanding path) in the right. Each path consists of six layers. The encoder module often reduces the input's resolution, and the decoder module struggles to produce fine-grained segmentations. Skip connections added from earlier layers enable the combination of fine layer features with coarse features and help ResU-Net-c make the

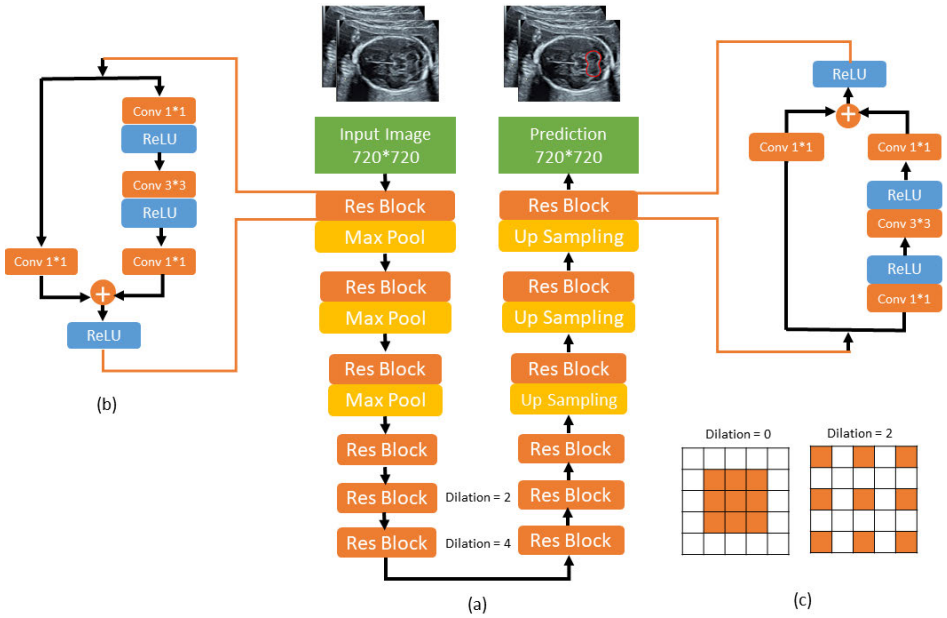

FIGURE 3. The proposed ResU-Net-c. (a) The network architecture. (b) Architecture of a single layer with residual connections. (c) Representation of 3 × 3 convolutional kernel with a dilation factor of 2.

TABLE 1. Architecture details of ResU-Net-c (Conv–convolution, ZP–zero padding, BN–batch normalization).

Residual Encoder Blocks	Layers in Encoder Block	Output size	Residual Decoder Blocks	Layers in Decoder Block	Output size
Contract Block 1	Conv →ZP →ReLU → BN →MaxPool2D	[64, 360, 360]	Expand Block 1	Conv →ZP →ReLU → BN →Negative ZP	[1, 720, 720]
Contract Block 2	Conv →ZP →ReLU → BN →MaxPool2D	[128, 180, 180]	Expand Block 2	Conv →ZP →ReLU → BN →UpSample2D	[64, 724, 724]
Contract Block 3	Conv →ZP →ReLU → BN →MaxPool2D	[256, 90, 90]	Expand Block 3	Conv →ZP →ReLU → BN →UpSample2D	[128, 362, 362]
Contract Block 4	Conv →ZP →ReLU → BN	[512, 90, 90]	Expand Block 4	Conv →ZP →ReLU → BN →UpSample2D	[256, 180, 180]
Dilate Contract Block 5	Dilate Conv →ZP → ReLU →BN | Dilation =2	[512, 90, 90]	Dilate Expand Block 5	Dilate Conv →ZP → ReLU →BN | Dilation =2	[512, 90, 90]
Dilate Contract Block 6	Dilate Conv →ZP → ReLU →BN | Dilation =4	[512, 90, 90]	Dilate Expand Block 6	Dilate Conv →ZP → ReLU →BN | Dilation =4	[512, 90, 90]

detection of the global structure during segmentation with fine-grain details. The residual blocks are incorporated into the network to provide the details necessary to reconstruct the cerebellum shape with appropriate boundaries. A layer built using a residual block is shown in Fig. 3(b). Unlike in standard U-Net architecture, the long skip connections between the corresponding feature maps of encoder and decoder modules are not used. We introduced short skip connections within the block to allow faster convergences when training and permit deeper models to be trained.

Two types of convolution operations (1×1 and 3×3) are performed in each residual block. After the convolutional operation, zero padding operations are used to keep the size of the feature maps unaltered. The number of feature maps successively increases from lower to higher layers, and we set the feature maps as 64, 128, 256, 512, 512, and 512 from layers 1 to 6, respectively. The detailed listing of each layer with it's output feature maps is shown in Table 1.

Dilated convolutions at the rate 2 and 4 are applied at layers 5 and 6, respectively. Dilated convolution broadens the receptive field with respect to a given constant filter size while preserving the full spatial dimension. Fig. 3(c) shows the representation of dilated convolution spaced according to the specified dilation rate. In the proposed ResU-Net-c architecture, the last two pooling layers are replaced with dilated convolutions to maintain the same field of view while preventing the loss of spatial information. ReLU is used as an activation function in all the layers to introduce non-linearity in the network. Max pooling operation with a stride of 2 pixels is used from 1-3 layers in the encoder part. This performs down sampling operation along the spatial dimensions and reduces the number of parameters and computations in the network to avoid overfitting. It also helps to provide a scale-invariant representation of the input image. Similarly, we used up-sampling layers in the decoder part to increase the spatial resolution of feature maps. The max pooling and up sampling layers do not have a skip connection between the input and the output. The output of the pooling and up-sampling layers are fed to their corresponding encoder and decoder blocks. The prediction layer consists of a single

convolution layer with kernel size 1×1. Softmax layer is used to predict the label map. We implemented ResU-Net-c in PyTorch using 12GB NVIDIA Tesla K80 GPU (NVIDIA, Santa Clara, CA, USA).

D. LOSS FUNCTIONS

A loss function is used to measure and minimize the difference between the predicted binary output and the binary ground truth. We used the Dice Loss (DL) [16], the combination of DL and Binary Cross-Entropy (BCE) [17], and Focal Tversky Loss (FTL) [18]. DSC was selected to measure the overlap of the segmented regions and BCE to quantify the pixel wise agreement between the output and the ground truth. FTL was used to solve the class imbalance problems in training segmentation models.

1) DICE LOSS (DL)

Dice Score Coefficient (DSC) is an overlap index that is widely used in medical image segmentation to assess segmentation performance. The two-class DSC variant for class c is given as,

$$DSC_c = \frac{\sum_{i=1}^{N} P_{ic} g_{ic} + k}{\sum_{i=1}^{N} P_{ic} + g_{ic} + k} \tag{2}$$

where, $g_{ic} \in 0, 1$ and $p_{ic} \in 0, 1$ represent the ground truth label and the predicted label, respectively for each class c. N denotes the total number of pixels in an image. The k provides numerical stability to prevent division by zero. The final DL is defined as a minimization of the overlap between the prediction and ground truth.

$$DL_c = \sum_{c} 1 - DSC_c \tag{3}$$

2) COMBO LOSS (CL)

We used a combination of BCE loss and DL [17]. For two-class problems, BCE loss function can be expressed as follows,

$$BCE(g, p) = -1/N \sum_{i=1}^{N} [g_i.log(p_i) + (1 - g_i).log(1 - p_i)] \tag{4}$$

The CL is computed as below,

$$CL = 0.5 * BCE + DL \tag{5}$$

The CL is parameterized by an individual weight factor $w \in [0, 1]$ for each loss. Here, w for BCE and DL are 0.5 and 1, respectively.

3) FOCAL TVERSKY LOSS (FTL)

In a highly imbalanced data and small foreground area, false negative (FN) detections need to be weighted higher than false positives (FP) to improve recall rate. Tversky similarity index (TI) [18] is a generalization of DSC score which allows for flexibility in balancing FP and FNs,

$$TI_c = \frac{\sum_{i=1}^{N} p_{ic} g_{ic} + k}{\sum_{i=1}^{N} p_{ic} g_{ic} + \alpha \sum_{i=1}^{N} p_{i\bar{c}} g_{ic} + \beta \sum_{i=1}^{N} p_{ic} g_{i\bar{c}} + k} \tag{6}$$

TABLE 2. Number of model parameters and epochs.

Model	No. of parameters	No. of epochs	
U-Net	31,903,043	DL	250
		CL	250
		FTL	200
Attention U-Net	34,876,033	DL	250
		CL	250
		FTL	250
UNet++	10,196,033	DL	150
		CL	100
		FTL	300
ResU-Net-c (w/o Residual blocks)	31,903,043	DL	200
		CL	225
		FTL	200
ResU-Net-c (w/o dilation)	17,640,722	DL	200
		CL	125
		FTL	100
ResU-Net-c	17,640,722	DL	127
		CL	235
		FTL	120

Here, p_{ic} represents the probability that pixel i belong to the cerebellum class c and $p_{i\bar{c}}$ is the probability that pixel i is of the background class, $\bar{c}$. Similar definition applies to g_{ic} and $g_{i\bar{c}}$. The hyper parameters α and β are set to improve the recall when there is a large class imbalance. In all the experiments, α and β were set to 0.3 and 0.7.

TI loss is obtained by minimizing $\sum_c 1 - TL_c$. Another issue with DL is that it struggles to segment small foreground, as they do not contribute to the loss significantly. To address this, we use FTL, parametrized by $\gamma \in [1, 3]$, to provide a trade-off between easy background and hard ROI training samples. In all our experiments, γ was set as $4/3$.

$$FTL_c = \sum_{c} (1 - TL_c)^{1/\gamma} \tag{7}$$

E. TRAINING PARAMETERS

The Adam optimizer was used as the optimization algorithm in all the models. It had better performance compared with other algorithms in all the training experiments. The model parameters were initialized with the random weights. The learning rate was set to 1×10^{-6}. The batch size of training and validating datasets was 2, as large batch size slows down the training speed. Table 2 shows the number of trainable parameters and number of epochs used in all the models.

F. EVALUATION METRICS

Evaluation Metrics play an important role in assessing the outcomes of segmentation models. In this work, we have analysed our results using DSC, Hausdorff Distance (HD), precision, and recall.

1) HAUSDORFF DISTANCE (HD)

HD [19] is a measure of segmentation error. HD is used to determine the degree of closeness between two images. HD is computed between the boundaries of the predicted (X) and ground truth (Y) segmentations and is given as below,

$$HD(X, Y) = max(hd(X, Y), hd(Y, X)) \tag{8}$$

TABLE 3. Evaluation of fetal cerebellum segmentation.

Models	Loss function	DSC	Precision	Recall	HD	p values	Processing time (sec) /image
U-Net	DL	0.7853	0.8623	0.7475	35.10	2.10E-18	0.41
	CL	0.8591	0.8774	0.8607	25.92	8.47E-08	0.39
	FTL	0.7877	0.7830	0.8334	42.10	6.93E-23	0.42
Attention U-Net	DL	0.8462	0.9232	0.8082	28.78	1.01E-03	0.17
	CL	0.8745	0.9448	0.8407	22.34	1.41E-03	0.13
	FTL	0.8462	0.9232	0.8082	29.30	3.22E-02	0.16
U-Net ++	DL	0.8353	0.8565	0.8467	31.28	4.14E-12	0.21
	CL	0.897	0.9518	0.8687	21.78	5.60E-03	0.20
	FTL	0.8692	0.9060	0.8694	25.26	3.10E-06	0.21
ResU-Net-c (w/o Residual Block)	DL	0.8739	0.8622	0.8990	25.23	1.08E-10	0.30
	CL	0.8760	0.8837	0.8828	24.28	8.31E-07	0.31
	FTL	0.8590	0.8427	0.8996	29.36	7.35E-12	0.34
ResU-Net-cu (w/o dilation)	DL	0.8492	0.9008	0.8259	29.17	8.41E-09	0.29
	CL	0.8530	0.9386	0.8069	32.42	3.98E-12	0.29
	FTL	0.8109	0.7664	0.8865	39.72	5.74E-08	0.29
ResU-Net-cu (Proposed)	**DL**	**0.9165**	0.9286	0.9163	**17.85**	—	0.30
	CL	0.9096	**0.9477**	0.8887	17.91	9.43E-01	0.32
	FTL	0.9105	0.8994	**0.9356**	19.25	3.04E-01	0.34

where,

$$h(X, Y) = maxmin|x - y| \tag{9}$$

where, $x \in X, y \in Y$

2) PRECISION AND RECALL

Precision and recall are measures of segmentation performance in terms of over and under segmentation. Low precision scores suggests over segmentation while low recall scores suggests under segmentation. True positive (TP) represents a pixel that is correctly predicted as ground truth. True negative (TN) represents a pixel that is correctly predicted as not belonging to ground truth. False Positive (FP) represents a pixel predicted incorrectly as ground truth. False Negative (FN) represents a pixel predicted incorrectly not as ground truth.

Precision is defined as,

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

Recall is defined as,

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

We compared our ResU-Net-c with well-established segmentation methods using U-Net as its base structure, such as U-Net [11], Attention U-Net [20], U-Net++ [17], and the proposed ResU-Net without dilation layers and without Residual Blocks. For ranking our models, we have chosen to use the DSC as the main evaluation parameter for comparison.

IV. RESULTS

Table 3 outlines the segmentation performance (for a single fold) of our ResU-Net-c against the comparison methods. The results of 5- fold cross validated training using DSC loss are shown in Table 4. ResU-Net-c achieved the highest mean DSC of 86.70%. Fig. 4 illustrates visual segmentation results of the proposed method and comparison methods, with original US images and delineated cerebellum overlaid on US images. Fig. 5 provides samples of false detections and original US images with ground truth labels. This sort of localization ability can help in better visualization of the fetal brain structures. Fig. 6 compares optimization learning curves for the proposed ResU-Net-c and comparison methods.

TABLE 4. Evaluation of fetal cerebellum segmentation using DL (mean ± standard deviation).

Model	DSC	Precision	Recall	HD	p value
U-Net	0.74±0.16	0.81±0.12	0.73±0.21	54.34±36.73	1.21E-52
Attention U-Net	0.81±0.14	0.87±0.11	0.79±0.19	38.84±29.76	1.07E-28
U-Net ++	0.81 ±0.15	0.85±0.11	0.80±0.19	41.14±32.41	5.97E-32
ResU-Net-C (w/o Res blocks)	0.83±0.12	0.85±0.10	0.82±0.16	35.90±26.02	1.97E-17
ResU-Net-C (w/o dilation)	0.77±0.17	0.89±0.10	0.72±0.21	50.30±37.82	1.41E-37
ResU-Net-C (Proposed)	**0.87±0.13**	**0.90±0.09**	**0.86±0.17**	**28.15±25.58**	

V. DISCUSSION

Table 3 shows that the proposed ResU-Net-c outperforms the comparison methods. This exemplifies the U-Net architecture's residual blocks effectiveness for US images. With the residual block, the low-level features from previous layers were directly combined with the high-level features from the recent layers, and this promotes the utilization of highly efficient features in the cerebellum segmentation.

Loss functions played an essential role in determining network performance. The performance metrics for all comparison methods with DL, CL, and FTL are shown in the Table 3. DL had better results in Residual U-Net-c, but the combined loss of DL and BCE, resulted in better performance among other comparison methods. Also, we investigated the combination loss of DL and BCE, to see if it can bring in different effects in the training process. We observed that the CL enforces a desired trade-off between false positives and negatives and avoids getting stuck in bad local minima as it leverages the DL [21]. The CL converges considerably faster than BCE during training. FTL has been shown to work well in highly imbalanced datasets, but it is observed that standard loss functions provide better optimization. We attribute this to the fact that FTL may not capture the uncertainties at boundaries. In practice, this results in segmentation maps with high precision but low recall. However, DL equally weighs FP and FN detections, and has shown better performance in the proposed method. Thus, the loss function's performance depends on the characteristics of the images used in training, such as skewness, and inconsistent boundaries.

Table 4 shows that the proposed method significantly outperformed the other comparison methods and had a higher DSC. The high accuracy of our method confirms that the use of residual blocks and dilated convolution, thus making it versatile to segment the cerebellum in US images.

Our experiments showed that our method segmented the cerebellum structure more accurately than other comparison methods (see Fig. 4). The contour obtained from the

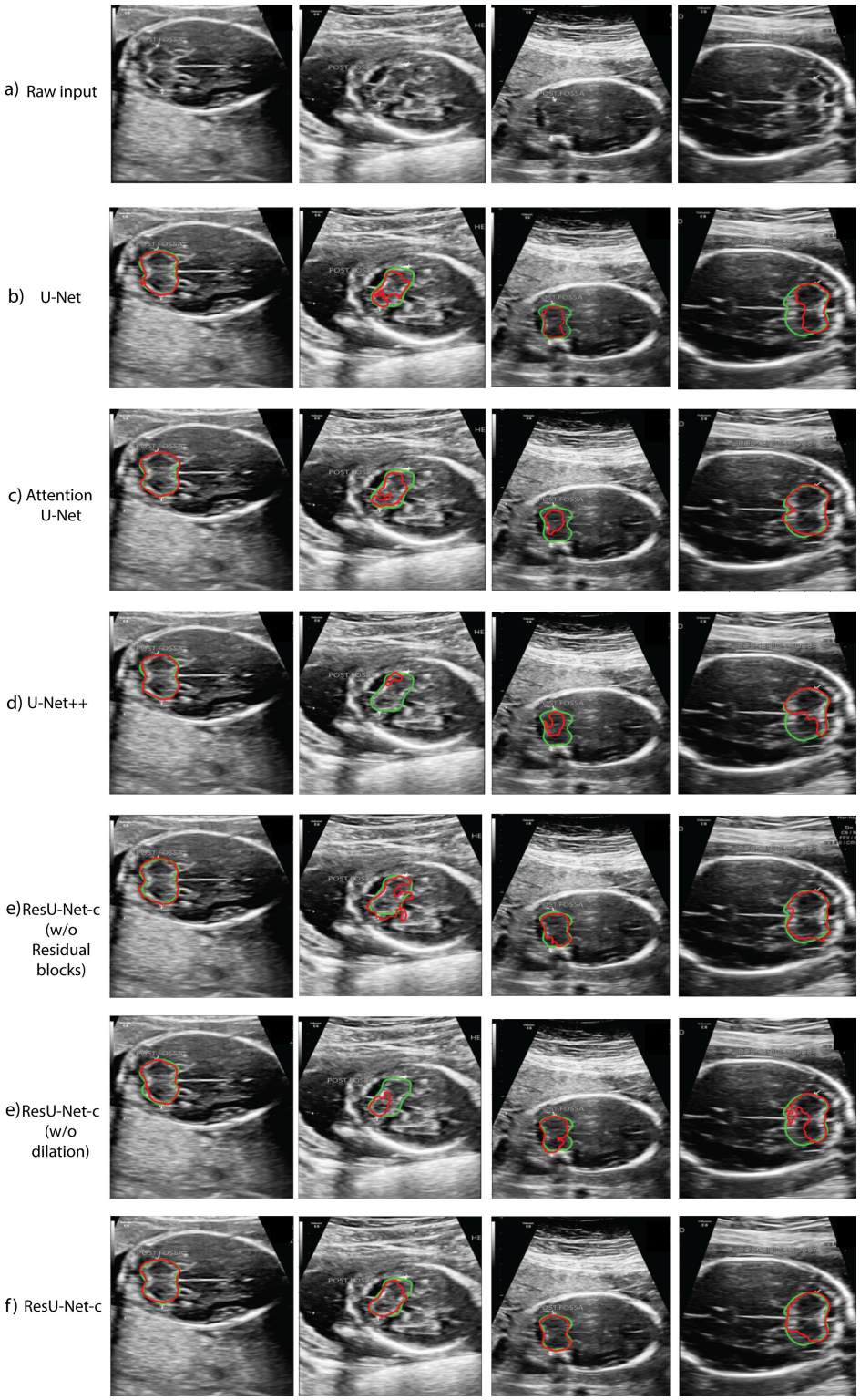

FIGURE 4. **Segmentation results. The rows depict the raw images followed by the segmentation results, and the columns are different US image samples. The ground truth labels are in green and the automated results are labeled in red.**

U-Net, U-Net++, Attention U-Net model did not depict the exact region of the cerebellum due to weak edges. They were plagued by contour leaks due to the lack of definite boundaries and did not cover the exact region of interest (see Fig. 4(d)-4th column). Applying the proposed method without the dilation layer resulted in segmented contours

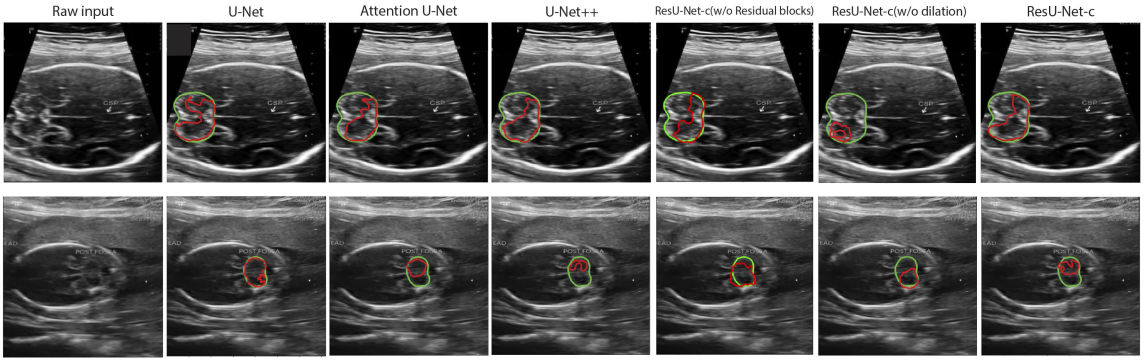

FIGURE 5. Samples of false predictions. The columns depict the raw images followed by the segmentation results, and the rows are different US image samples. The ground truth labels are in green and the automated results are labelled in red.

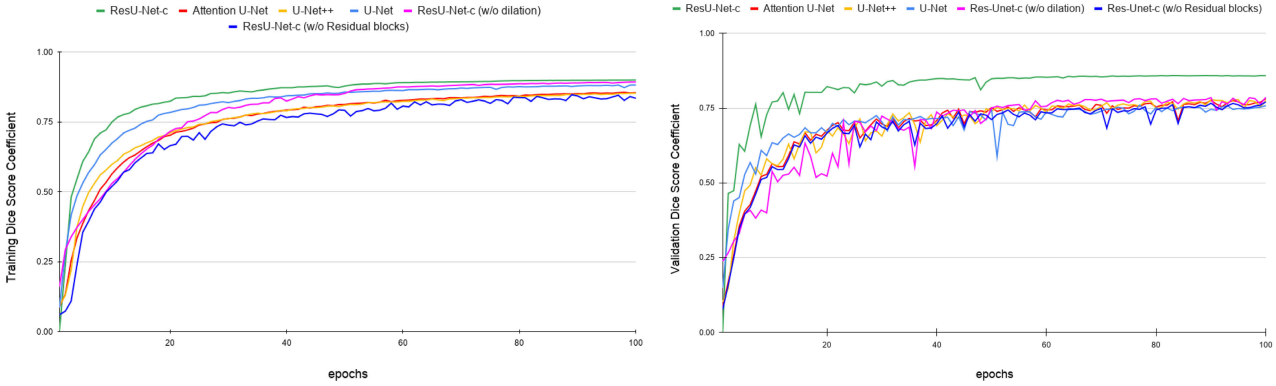

FIGURE 6. Learning curves (DSC) for the comparison method and proposed ResU-Net-c.

that crumbled inside the cerebellum. The boundaries of the cerebellum were not captured consistently due to poor contrast in the boundaries of the cerebellum (see Fig. 4(e)). Similar findings were observed in other comparison methods, suggesting that the use of dilation convolution is important for cerebellum segmentation. In all the samples, better visual results were obtained for the proposed ResU-Net-c.

These visual outcomes in Fig. 4 and Fig. 5 correspond to our quantitative findings in Table 3. ResU-Net-c achieved the highest DSC with DL function. The low precision value of the comparison methods indicates a large number of non-cerebellum pixels within the segmented contour. The higher precision of ResU-Net-c confirms that there are fewer FP with the predicted cerebellum. The superiority of the proposed algorithm over the other comparison methods is statistically significant ($p < 0.001$) and is demonstrated by comparing the HD values of the proposed method with other comparison methods using two-tailed paired t-test.

We observed that segmentation performance is poor in Fig. 5 for images where the boundaries of the fetal head is not fully visible and/or there is discontinuity across all methods. One of the reason for this low performance is attributed to the visualization of the structure of the skull in the fetal head. We recommend that such images may not be appropriate to be included in the automated processing of retrospective US images and requires image quality assessment.

The performance of ResU-Net-c (w/o Res Block) and ResU-Net-c (w/o dilation) is higher than the U-Net and lower than the proposed method. We attribute the increased performance of our method to the enriched semantic features learned through skip connections present in the residual blocks. The residual operations significantly improve the training and the testing performance without increasing the number of network parameters [22].

The ResU-Net-c (w/o dilation) results were lower across all measures, showing the importance of the inclusion of the dilation layer in the proposed method. Dilated convolution filters increase the receptive field of CNN without introducing additional parameters, which avoids the overfitting issues in the training process. Dilated convolutions used in our method supports the exponential expansion of the receptive field without loss of spatial resolution [23].

Fig. 6 shows that the proposed method has a steep acceleration in the learning curve than other methods. These quantitative findings demonstrate the accuracy, robustness, and reliability of the proposed method in the cerebellum segmentation.

With the use of dilation layer and residual blocks, the number of parameters needed to train the model has been reduced. ResU-Net-c has 17, 640, 722 trainable parameters, as shown in Table 2. The number of model parameters is lower for UNet++. Likewise, ResU-Net-c needed a lesser number of

epochs for convergence. The average processing times to test an image using our method and other comparison methods are shown in Table 3. The processing time of our method was lower compared to the U-Net but higher than Attenuation U-Net. Optimization of running time was not within the original scope of this research and has been left for future work.

Manual cerebellum measurements are easy to perform, and semi-automatic techniques are fast; however, these approaches all rely on user inputs and therefore are subject to be compromised in robustness and consistency. Further manual approaches are time-consuming when a large number of images are needed. Our automated segmentation method enables retrospective study that involves a large number of US Images. Our future work will involve image quality assessment as a prior to automated fetal US image segmentation.

Future work will also include the extension of our algorithm to prospective cerebellum measurements and quantification. We suggest that our method can potentially help to decrease operator dependency in clinical applications for the assessment of fetal health, thereby increasing robustness and reproducibility.

VI. CONCLUSION

We proposed a new semantic segmentation method to segment the cerebellum from 2D US images with a U-Net architecture combined with residual blocks. All the comparison models have been evaluated using three different loss functions: DL, FTL and a CL. The experimental results demonstrate that the proposed ResU-Net-c model outperformed in the segmentation tasks compared to the existing methods. This method can be extended to perform semantic segmentation of other fetal brain structures with biometric measurements.

ACKNOWLEDGMENT

(Vishal Singh and Pradeeba Sridar contributed equally to this work.)

REFERENCES

[1] E. Kelly, F. Meng, H. Fujita, F. Morgado, Y. Kazemi, L. C. Rice, C. Ren, C. O. Escamilla, J. M. Gibson, S. Sajadi, and R. J. Pendry, "Regulation of autism-relevant behaviors by cerebellar–prefrontal cortical circuits," *Nature Neurosci.*, vol. 23, no. 9, pp. 1102–1110, 2020.

[2] M. E. van der Heijden, J. S. Gill, and R. V. Sillitoe, "Abnormal cerebellar development in autism spectrum disorders," *Develop. Neurosci.*, pp. 1–10, Apr. 2021, doi: 10.1159/000515189.

[3] A. J. Spittle, L. W. Doyle, P. J. Anderson, T. E. Inder, K. J. Lee, R. N. Boyd, and J. L. Y. Cheong, "Reduced cerebellar diameter in very preterm infants with abnormal general movements," *Early Hum. Develop.*, vol. 86, no. 1, pp. 1–5, Jan. 2010.

[4] H. W. Park, H.-K. Yoon, S. B. Han, B. S. Lee, I. Y. Sung, K. S. Kim, and E. A. Kim, "Brain MRI measurements at a term-equivalent age and their relationship to neurodevelopmental outcomes," *Amer. J. Neuroradiol.*, vol. 35, no. 3, pp. 599–603, Mar. 2014.

[5] M. An, "Unusual brain growth patterns in early life in patients with autistic disorder," *Neurology*, vol. 57, pp. 245–254, Jul. 2001, doi: 10.1212/WNL.57.2.245.

[6] A. V. Shevelkin, C. Ihenatu, and M. V. Pletnikov, "Pre-clinical models of neurodevelopmental disorders: Focus on the cerebellum," *Rev. Neurosci.*, vol. 25, no. 2, pp. 177–194, Jan. 2014.

[7] X. Liu, J. Yu, Y. Wang, and P. Chen, "Automatic localization of the fetal cerebellum on 3D ultrasound volumes," *Med. Phys.*, vol. 40, no. 11, 2013, Art. no. 112902.

[8] M. Yaqub, R. Cuingnet, R. Napolitano, D. Roundhill, A. Papageorghiou, R. Ardon, and J. A. Noble, "Volumetric segmentation of key fetal brain structures in 3D ultrasound," in *Proc. Int. Workshop Mach. Learn. Med. Imag.* Cham, Switzerland: Springer, 2013, pp. 25–32.

[9] M. E. Roy-Lacroix, F. Moretti, Z. M. Ferraro, L. Brosseau, J. Clancy, and K. Fung-Kee-Fung, "A comparison of standard two-dimensional ultrasound to three-dimensional volume sonography for routine second-trimester fetal imaging," *J. Perinatol.*, vol. 37, no. 4, pp. 380–386, Apr. 2017.

[10] M. R. López, F. A. Cosío, B. E. Ramírez, and J. O. Montiel, "Shape model and Hermite features for the segmentation of the cerebellum in fetal ultrasound," in *Proc. 14th Int. Symp. Med. Inf. Process. Anal.*, Dec. 2018, Art. no. 1097514.

[11] 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.* Cham, Switzerland: Springer, 2015, pp. 234–241.

[12] L. Venturini, A. T. Papageorghiou, J. A. Noble, and A. I. Namburete, "Multi-task CNN for structural semantic segmentation in 3D fetal brain ultrasound," in *Proc. Annu. Conf. Med. Image Understand. Anal.* Cham, Switzerland: Springer, 2019, pp. 164–173.

[13] R. Ramos, J. Olveres, B. Escalante-Ramírez, and F. A. Cosío, "Deep learning approach for cerebellum localization in prenatal ultrasound images," *Proc. SPIE*, vol. 11353, Apr. 2020, Art. no. 1135322.

[14] 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.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.

[16] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.

[17] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham, Switzerland: Springer, 2018, pp. 3–11.

[18] N. Abraham and N. M. Khan, "A novel focal tversky loss function with improved attention U-Net for lesion segmentation," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 683–687.

[19] D. Karimi and S. E. Salcudean, "Reducing the hausdorff distance in medical image segmentation with convolutional neural networks," *IEEE Trans. Med. Imag.*, vol. 39, no. 2, pp. 499–513, Feb. 2020.

[20] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Rueckert, "Attention gated networks: Learning to leverage salient regions in medical images," *Med. Image Anal.*, vol. 53, pp. 197–207, Apr. 2019.

[21] S. A. Taghanaki, Y. Zheng, S. K. Zhou, B. Georgescu, P. Sharma, D. Xu, D. Comaniciu, and G. Hamarneh, "Combo loss: Handling input and output imbalance in multi-organ segmentation," *Comput. Med. Imag. Graph.*, vol. 75, pp. 24–33, Jul. 2019.

[22] G. A. Francia, C. Pedraza, M. Aceves, and S. Tovar-Arriaga, "Chaining a U-Net with a residual U-Net for retinal blood vessels segmentation," *IEEE Access*, vol. 8, pp. 38493–38500, 2020.

[23] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *Proc. 4th Int. Conf. Learn. Represent. (ICLR)*, 2016, pp. 1–13.

VISHAL SINGH is currently pursuing a dual Degree Program (B.Tech and M.Tech) in biomedical engineering with the Department of Engineering Design, IIT Madras, Chennai, India. He has done a research internship in his junior year at The University of Sydney, Australia, and a Research Engineer with Philips Healthcare, Bengaluru, India, in his senior year. His research interest includes medical imaging using deep learning methods.

PRADEEBA SRIDAR received the bachelor's degree in electronics and communication engineering from Anna University, Chennai, India, in 2009, the master's degree in communication engineering from VIT University, Vellore, India, in 2011, and the Ph.D. degree in the area of medical image analysis from the Department of Engineering Design, IIT Madras (IIT-M), Chennai. During her Ph.D. degree, she conducted research in fetal ultrasound image analysis in collaboration with researchers from IIT-M, The University of Sydney, Australia, and the Nepean Hospital, Australia. She is currently an Honorary Associate with the Sydney Medical School Nepean. Her research interests include the development of an automated framework for fetal ultrasound image analysis, translational engineering, and machine learning for medical image analysis. She was a recipient of the prestigious Endeavour Research Fellowship, in 2014.

JINMAN KIM (Member, IEEE) received the Ph.D. degree in computer science (biomedical image analysis) from The University of Sydney, in 2006. He was an ARC Postdoctoral Research Fellow with The University of Sydney and then a Marie Curie Senior Research Fellow with the University of Geneva, prior to joining The University of Sydney, in 2013, as a Faculty Member. His research interests include image processing and visualization of multi-modal biomedical imaging data.

RALPH NANAN received the Ph.D. degree in basic immunology from the Children's University Hospital, Wuerzburg, Germany, in 1991, and the Habilitation degree in clinical immunology, in 2002. He is currently the Chair and a Professor of paediatrics with the Sydney Medical School Nepean and the Sub-Dean of research. He is also the Director of the Charles Perkins Centre Nepean and leads the Developmental Origins of Health and Disease Project Node. He started his scientific career in paediatrics at the Children's University Hospital. He accept the position of the Director of paediatrics with the Nepean Hospital, in 2003, and then his current position, in 2005. Since beginning at Nepean, he has established the Paediatric Immunology Laboratory and developed a comprehensive paediatric allergy service. He has also been heavily involved in policy development on a state level. He is also the Chief Investigator on several clinical cohort studies. Apart from this, he has multiple teaching and supervisory responsibilities at The University of Sydney and regularly speaks at national and international conferences.

N. POORNIMA received the M.B.B.S. and D.N.B. (RD) degrees. She was specialized in diagnostic radiology from the premier institute of Barnard Institute of Radiology, Madras Medical College, Chennai, India. She has a brilliant academic record as a Postgraduate Student, and joined Precision Diagnostics as a Junior Consultant, at the Fortis Malar Hospital. She is experienced and skillful in interventional procedures. Her special research interest includes gastrointestinal and general radiology. She has been certified by the Fetal Medicine Foundation, U.K., to perform advanced obstetric scans.

SHANMUGA PRIYA received the M.B.B.S., D.M.R.D., and D.N.B. (RD) degrees. She received the master's degree from the reputed Sri Ramachandra Medical College, Chennai, India. She was a Teaching Faculty with the Sri Ramachandra Medical College, before joining Precision Diagnostics. She has had special training in fetal imaging and fetal echo cadiolography. She has also been certified by the Fetal Medicine Foundation, U.K., to perform advanced obstetric scans. She is also adept at performing interventional gynecological procedures, such as ultrasound guided salpingography, saline infusion studies, and hysterosalpingography.

G. SAMEERA REDDY graduated from the Sri Ramachandra Medical College, Chennai, India. She received the M.B.B.S. and D.M.R.D. degrees. She has served as a Faculty Member with the Sri Ramachandra Medical College on the teaching side, before she went to the University of San Diego, CA, to train under the illustrious Prof. Resnick in musculoskeletal imaging. She had a two year stint at the Apollo Heart Centre, Chennai, conducting and interpreting CT coronary angiograms. She has been certified by the Fetal Medicine Foundation, U.K., to perform advanced obstetric scans.

SATHYABAMA CHANDRASEKARAN received the M.B.B.S. degree from the Jawaharlal Institute of Postgraduate Medical Education and Research (JIPMER) and the D.M.R.D. and M.D. degrees from the Madras Medical College, Chennai, India. She had her initial training as a Resident Radiologist under the guidance of the great doyen of Radiology, Prof. Arcot Gajaraj, in Apollo Hospitals. She then went on to be a part of Precision Diagnostics, a Reputed Diagnostic Group, of which she was the Director. She has published many national and international articles on abdominal doppler studies like portal hypertension in adults and children and has worked with Prof. Soloman Victor and Prof. V. Jayanthi to do extensive research on BuddChiari syndrome. Her work has been recognized in international forums. She has served as the Head of the Department of Radiology and Imaging, Fortis Malar Hospitals, for ten years, before launching Athena.

RAMARATHNAM KRISHNAKUMAR is currently an Institute Professor and the Perry L. Blackshear Chair of IIT Madras, Chennai, India. His research interests include image processing, biomedical engineering, and non-linear mechanics. He is a Fellow of the Indian National Academy of Engineers.