IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 26 September 2022, accepted 14 October 2022, date of publication 19 October 2022, date of current version 4 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3215973

RESEARCH ARTICLE

Echocardiographic Image Segmentation for Diagnosing Fetal Cardiac Rhabdomyoma During Pregnancy Using Deep Learning

SUDHAKAR SENGAN[1], (Member, IEEE), ABOLFAZL MEHBODNIYA[2], (Senior Member, IEEE), SURBHI BHATIA[3], (Member, IEEE), S. S. SARANYA[4], (Member, IEEE), MESHAL ALHARBI[5], SHAKILA BASHEER[6], AND V. SUBRAMANIYASWAMY[7]

[1]Department of Computer Science and Engineering, PSN College of Engineering and Technology, Tirunelveli, Tamil Nadu 627152, India
[2]Department of Electronics and Communications Engineering, Kuwait College of Science and Technology, Safat 20185145, Kuwait
[3]Department of Information Systems, College of Computer Science and Information Technology, King Faisal University, Al Hofuf 36362, Saudi Arabia
[4]Department of Computer Science and Engineering, School of Computing, SRM Institute of Science and Technology, Kattankulathur Campus, Chennai, Tamil Nadu 603203, India
[5]Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia
[6]Department of Information Systems, College of Computer and Information Science, Princess Nourah Bint Abdulrahman University, Riyadh 11671, Saudi Arabia
[7]School of Computing, SASTRA Deemed University, Thanjavur, Tamil Nadu 613401, India

Corresponding author: Sudhakar Sengan (sudhasengan@gmail.com)

This work was supported by Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia, through the Researchers Supporting Project PNURSP2022R195.

ABSTRACT Automated interpretation of cardiac images has the potential to change clinical practice in many ways. For example, it could make it possible for non-experts in primary care and rural settings to test the heart's function over time. In this paper, we tested the research hypothesis that recent developments in computer vision would make it possible to create a fully automated, scalable analysis for echocardiogram interpretation, covering all of the steps from view identification and Medical Image Segmentation (MIS) to structure and function quantification and Fetal Cardiac RhabDomyomas (FCRD) detection. Even though they are rare, FCRDs are the most frequent cause of Fetal Cardiac Tumor (FCT). When it comes to diagnosing and monitoring fetuses with an injured circulatory system, imaging (particularly echocardiography (ECG)) has proven helpful in the field of fetal cardiology. Because of the severe lack of qualified and experienced sonographers, it is very challenging to diagnose Cardiac RhabDomyomas (CRD). Prior to delivery, accurate segmentation of the FC to identify structural cardiac defects is critical for minimizing the illness among newborns. To automate the process of segmenting the cardiac chamber for the CRD, we propose a novel Attention-Residual Network-based V-Net architecture (ARVNet). In this study, examinations were performed on Fetal Rhabdomyomas noted in the Right Ventricle (FRRV), Fetal Rhabdomyomas noted in the Left Ventricle (FRLV), Fetal Rhabdomyomas noted in the Right Atrium (FRRA), Fetal Rhabdomyomas noted in the Left Atrium (FRLA), Fetal Rhabdomyomas noted in the Tricuspid Valve (FRTV). Images without Rhabdomyoma mean "Normal Condition (NC)" at Selvam Hospital in Melapalayam, Tirunelveli, Tamil Nadu, India. Even with a relatively small number of datasets, the proposed technique possesses high CRD detection performance, as evidenced by the results. The results showed that the proposed model did a good job segmenting all the views, with a specificity of 99.7% and a Dice coefficient similarity of 99.8%. It also did well at finding CRDs, with an average mean accuracy of around 99.85%.

INDEX TERMS Fetal cardiac tumor, deep learning, cardiac rhabdomyomas, v-net, accuracy, specificity, similarity.

The associate editor coordinating the review of this manuscript and approving it for publication was Md. Kafiul Islam.

I. INTRODUCTION

Congenital heart abnormalities comprise 2.8% of FCTs, and the majority of those are benign and uncommon. 60% of FCT

cases is rhabdomyomas, the most prevalent kind. Usually discovered 20 weeks after delivery, Fetal Cardiac RhabDomyomas (FCRD) frequently develop in the ventricular cavity [1]. The prognosis depends on the size of the tumour and whether there are any additional issues; an isolated CRD typically predicts a better prognosis. Your child's heart may develop CRD on either side of the chamber (Right or Left). Usually, they are in the Right or Left ventricle (lower chambers of the heart). They can occasionally be seen in the walls separating the ventricles or the atria (upper chambers) [2], [3]. The tumours can range from a few millimetres (*mm*) to several centimetres (*cm*) in size. When observed with the naked eye, they seem to be Yellow or White. CRD is more common in babies and kids with tuberous sclerosis. Tuberous sclerosis, which accounts for 50% to 60% of cases, is present in at least 8 out of 10 CRD patients. The autosomal dominant multisystem disorder Tuberous Sclerosis Complex (TSC) is defined by hamartomas in various systems of organs, including the skin, brain, kidney, lungs, and heart. Complications from these changes may include learning disabilities, epilepsy, behavioural issues, and renal failure. In neonates, the incidence of TSC varies from 1/6000 to 1/1000; in adults, it is 1/8000 [4]. A mutation in either the TSC1 or TSC2 gene results [5].

Prenatal screening can identify the earliest signs of tuberous sclerosis, which typically affects the heart and brain. Rhabdomyomas is just an early symptom and could be the only one present before the further development of clinical symptoms. This is why in cases of CRD, the diagnosis of tuberous sclerosis must be taken into account. According to a meta-analysis by [6], only 13.7% of cases of CRD were discovered prior to the 24^{th} gestational week, while the majority of cases are discovered beyond that point. The 2012 International Tuberous Sclerosis Complex Consensus Guidelines noted that genetic diagnosis can be utilised as an autonomous diagnostic criterion and added it to the diagnostic criteria for TSC. In fetuses and newborns, CRD could be the only symptom of TSC [7]. Thus, if isolated CRD is identified during prenatal testing, doctors should also perform further TSC-related gene testing to ascertain whether the fetus has TSC. They should also take into account the space-occupying effect of the tumour itself and hemodynamic abnormalities.

Imaging tests are used to diagnose CRDs. These examinations might happen before or after childbirth. Prenatal ultrasonography is frequently used to detect a tumour for the first time. During weeks 20 to 30 of pregnancy, the ultrasound shows signs of the tumour. Another method of diagnosis is Fetal Echocardiography (FECG). Until a baby is born, this type of ultrasound looks explicitly for cardiac issues. The heart's anatomy can be sufficiently shown by Transthoracic Echocardiography (TTECG), which can also be used to assess hemodynamics as well as cardiorespiratory performance. It is now the most frequently used non-invasive investigation procedure for CRDs. The heterogeneity of ECG diagnosis is caused by various causes, including the heart's natural pulse variability, speckled noise and artefacts, and variations in typical ECG views within and between classes [8]. Accurate diagnosis of TTECG is complex and requires minimum time consumption, and it severely depends on the accurate interpretation of each ECG view by professional heart specialists. Anatomical structures, as well as the spatial configuration of CRDs, seem to be complicated and unpredictable, and exact finding across TTECG is challenging and require less time computation.

Due to the high number of paediatricians in India, there is a prodigious requirement for experts in ECG for the diagnosis of CRD. From various angles, the operator must gather video streams and examine the morphology of tissues and organs in order to make an ECG diagnosis. The operator's individual technical abilities have a significant impact on accurate diagnosis. Most primary hospitals that lack experienced ECG operators find it challenging to effectively identify CRDs because of the extensive training required to become an operator in ECG [9]. In order to accurately and timely diagnose CRDs and help ECG operators avoid misdiagnosis brought on by unnatural causes, an autonomous diagnostic method based upon ECG analysis is urgently needed.

The field of Computer Vision (CV), which includes automated image interpretation, is a subset of Machine Learning (ML), which is the study of how computers can replicate human vision. Despite the fact that CV has long been used in medical imaging, new developments in CV algorithms, processing power, as well as a considerable increase in digitally labelled data have led to a successful comprehension in classification performance for a number of test cases, particularly retinal and skin illness. But there are more difficulties with ECG than these. A regular ECG comprises more than 70 recordings gathered from various angles instead of just one still image, and the angles are not labelled in every study [10]. The intrinsic beat-to-beat variability within heart function and the variability from the process of estimating a 3-D object utilizing 2-D cross-sectional images could all cause measurements to differ from one video to the next. We expected that this ECG could benefit from an automated learning technique to support human interpretation because of the magnitude of this variability and the vast amount of multi-dimensional information within every study that frequently goes unused.

Convolutional Neural Networks (CNN) based on DL are an AI strategy that can be used to diagnose prenatal objects [11], [12], [13]. Regarding CNNs' capability in segmentation, classification, as well as detection regarding medical imaging, several researchers have achieved strong results [14], [15]. CNN's programmes that carry out adaptation tasks without explicit programming learn through data and produce precise forecasts or judgments based on prior information [16]. However, leakage through missing boundaries produced by intra-chamber walls is still an issue in the FECG studies based on CNN's [17]. By choosing appropriate Regions of Interest (RoI), prior studies suggested extracting patterns of cardiac anatomy.

Moreover, inside the testing, the surveyed object detection methods successfully detected only one candidate region;

multiple candidate detection is challenging to implement. A classification strategy has been used to resolve this problem [18]. However, in this instance, segmentation, classification, and object identification are all made independently; in other words, such methods are not done instantaneously. Instead, CNNs are used for learning just one job at a time. For FC imaging, multi-task learning is crucial because it allows a model to segment multiple regions, choose multiple patients, classify numerous RoI, and detect multiple diagnostic items.

For images that is related to both medical and non-medical fields, many models based on deep convolutional segmentation with encoder-decoder formatted networks have been presented. These models include the 2D U-Net [19], 3D U-Net [20], 3D V-Net [21], and the 2D SegNet [22]. The issues related to the segmentation of binary and multi-class medical images have seen extensive use of these models, with or without changes. One must deal with two different forms of imbalance difficulties while performing segmentation, particularly when employing DL networks: (a) Input imbalance but rather inter-class imbalance throughout training, that is, during segmentation, substantially a few foreground voxels/pixels relative to the numerous background voxels, as well as tiny objects or classes associated with particular other more oversized classes/objects and the background in multi-class segmentation. As a result, classes with much more observations (*i.e.,* voxels) hide those with fewer observations. (b) An imbalanced output, with False Positives (FP) as well as False Negatives (FN) in inference, are inevitable. The background voxels that are mistakenly identified as the target object are FPs (or different objects in the event of multi-class). In the multi-organ segmentation context, the voxels of such a target item, which is incorrectly identified as contextual or as a different organ, is known as FN. Clearly, the ideal situation is to eradicate both FP and FN completely. But in real-world systems, one grows as the other shrinks. Lowering the False Positive Rate (FPR) is more crucial for some applications than minimizing the False Negative Rate (FNR) or vice versa. The suitable formulation of a Dice-Based Loss Function (DBLF) is essential when DL is performed on imbalanced data. U-Net (2D and 3D) and 2D SegNet apply a DBLF while simulating ground truth segmentation masks for such a test image, and 3D V-Net minimises Cross-Entropy (CE) loss. Although it may be able to regulate the FP and FN output imbalances, it performs poorly when segmenting images with many input class imbalances. A DBLF must have demonstrated higher performance in binary-class segmentation issues among all strategies suggested for addressing an input imbalance problem [23]. However, earlier works have not examined the DBLF's capacity to regulate the trade-off among FP as well as FN (*i.e.,* output imbalance). It is difficult for a traditional Dice optimization algorithm to properly manage the trade-offs for certain types of medical image processing.

Classification, segmentation, object detection and recognition, and regression are just a few of the steps that Machine Learning (ML), Deep Learning (DL), Artificial Neural Networks (ANN), and Reinforcement Learning (RL) are applied to perform on medical data. [24], [25], [26]. The majority of medical imaging data, including that from X-ray, fluoroscopy, MRI, ultrasound or CT scanners, computed tomography, elastography, sensory imaging, and infrared thermometers, has been subject to AI-automated diagnosis using Computer-Aided Detection (CAD) [27], [28], [29]. Digitized medical images, on the other hand, introduce them to a wealth of new personal data, options, and issues. As a result, methodologies are able to address some of these difficulties by identifying imaging anomalies with remarkable sensitivity and accuracy. With the potential to improve disease diagnoses, these techniques promise to improve tissue-based identification and classification [30].

Currently, the use of artificial intelligence algorithms in medical data has been discussed extensively across a wide range of medical fields of study, including classifying cardiovascular anomalies, finding broken bones and other neurological problems, supporting the finding of neurodegenerative disorders, minimising thoracic health problems and factors, screening for common tumors, and many other treatment plans and diagnosis actions [31], [32], [33], [34]. Moreover, AI methods have demonstrated the ability to produce feasible findings when applied to premature medical data, such as assessing fetal growth at each stage of pregnancy, inferring the health of an early pregnancy birth canal, and evaluating possible health problems [35], [36], [37], [38], [39]. Several fetal illnesses and adverse pregnancy results in complex etiologies and pathogeneses, such as amniotic band disorder, congenital diaphragmatic burst, congenital high airway meddling syndrome, fetal bowel obstruction, gastroschisis, omphalocele, pulmonary sequestration, and sacrococcygeal malignant tumor, may be performed to detect using Artificial intelligence techniques. [40], [41], [42]. In order to avoid and decrease unfavourable conclusions, provide consideration of fetal irregularities and illnesses throughout pregnancy, and severely reduce the need for more invasive problem-solving measures that may be harmful to the fetus, additional research is required [43].

The use of AI techniques during pregnancy is possible with a distinct set of imaging technologies, such as ultrasound, MRI, and Computed Tomography (CT) [44]. Ultrasound imaging has emerged in clinical studies and is used throughout every trimester of pregnancy. Diagnosis and monitoring of fetal growth and progress both rely heavily on ultrasound technology. Furthermore, ultrasound can provide clear fetal anatomical evidence, high-quality medical images, and increased investigative precision [45], [37]. The use of ultrasound has numerous advantages and few limitations. Researchers reviewed and presented a number of methods based on model-driven Medical Image Segmentation (MIS) prior to the widespread adoption of DL. Several model-driven methods for medical image analysis, such as image clustering, region growing, and Random Forest (RF), were thoroughly summarised by [46]. According to various mathematical models, the author summarised various segmentation models on medical images. Data-driven

studies have become increasingly more common, while model-driven studies have become increasingly rare, in recent years, for MIS. In [47], the author gave a comprehensive summary of how DL has been used for analysing medical images. Medical image recording, human-anatomy and cell structure detection, tissue segmentation, CAD, and prognosis are just some of the areas where this review highlights the development of ML and DL. Image classification, object detection, segmentation, registration, and other basic functions are all included in the research study of DL methods that was recently reported [48].

Analysis has divided DL-based MIS solutions into six groups: deep architectural, data synthesis-based, loss function-based, sequenced models, weakly supervised, and multi-task methods. Recently, [49] discussed the development of semantics and MIS. To develop a complete survey on MIS, [50] reviewed classical ML algorithms such as Markov random fields, k-means clustering, and RF, and reviewed the latest DL models such as ANNs, CNNs, and RNNs. In [51], we looked at how to deal with MIS despite having imperfect data sets, which can be hampered by both a lack of annotations and low-quality annotations. Each and every one of these samples contributes significantly to the growth of MIS methods. Three elements of approaches (network structures), training methods, and challenges were reviewed in [52]. The primary, general network models used for MIS are explained in the segment on network models. The J-DIGIT deep learning model, which is used to train deep neural network models, is supported in the section on training techniques. The difficulties of implementing DL strategies in MIS are discussed in detail. The developments in the use of, or possible future use of, DL in radiotherapy were analysed in [53]. Current DL-based segmentation methods for quantitative brain MRI images were reviewed in [54]. Three common forms of poor supervision were emphasised by [55]: oversight that was insufficient, imprecise, and inaccurate. Primarily on Dice scores or Jaccard indices, [56] concentrates on analysing and summarising the optimization techniques applied to MIS tasks.

Using standard 12-lead ECGs, [57] presented a novel ECG diagnostic system for the detection of 13 distinct diseases. To achieve the accuracy of a human diagnostician's task at the speed of an automated method, the proposed model recommends a set of rules. A specialist's input and a database of 284,000 ECGs served as the foundation for the rules' design, which was completed. Professionally validating the system, it was found to have a trustworthiness of 80.8%. For the prognosis of ARR, Normal Sinus Rhythm (NSR), and CHF, [58] presents another innovative ensemble method based on Shifted One-Dimensional Local Binary Patterns (S-1D-LBP) and LSTM. According to the common decisions of LSTM models with nine distinct input signals, the ECG signals are classified. The proposed approach achieved a high (99.6%) success rate when tested using ECG signals for the method (ARR, NSR, and CHF). A new residual block for CNNs for MIS was proposed by [59]. FocusNet++ is a group attention mechanism created by combining attention mechanisms with group convolution layers. In addition, they used a hybrid loss based on balanced cross-entropy, Tversky loss, and adaptive logarithmic loss to improve performance and achieve rapid convergence. In terms of various benchmark metrics for the ISIC 2018 melanoma segmentation, their findings demonstrate that FocusNet++ achieves cutting-edge results.

Despite a growing body of research on the potential benefits of using AI methodologies for ultrasound medical imaging diagnosis and treatment of pregnancy complications, little attention has been paid to the problem of identifying CRDs in US images acquired during pregnancy, which is what this research attempts to cure. Presently, there is considerable interest in the purpose of DL technology for medical imaging. Many researchers are now interested in the problem of how to identify and segment the lesions in medical images robotically. This issue was addressed by [60]'s U-Net proposal at the MICCAI Summit in 2015, which was a major development for DL in MIS. The encoder, bottleneck module, and decoder make up U-Net, which is a Fully Convolutional Network (FCN) applied to MIS. U-U-shaped Net's model, when coupled with a background relevant data, fast training speed, and a small amount of data, makes it suitable person for MIS.

A new Attention Gate (AG) model was put forth by [61] for the analysis of medical imaging. The model trained with AG evolves to restrict absolutely meaningless regions of an input image and highlight significant characteristics suitable for particular tasks in an indirect manner. The application of overt exterior tissue and organ localization units of cascading CNNs is eliminated as a result, which is helpful. A standard CNN model, such as U-Net, could be combined with AG to improve the model's sensitivity and accuracy. The Context Encoder Network (CE-Net), which uses pre-trained Res-Net blocks as fixed feature extractors, was proposed by [62] as a way to obtain more sophisticated data and retain spatial data for 2D segmentation. The feature encoder, context extractor, and feature decoder make up the majority of its three components. A recently developed block called Dense Atrous Convolution (DAC) and a block called Residual Multikernel Pooling (RMP) help compensate the context extractor. In comparison to the original U-Net approach, the recently developed CE-Net is widely used for segmentation in 2D medical imaging. UNet++, a new and improved neural network model for MIS, was suggested by [63] to develop the segmentation process.

Later, a new version of UNET 3+ was developed by [64] to increase accuracy, particularly for organs of various sizes. In order to learn a visual image from a set of aggregated full-scale feature maps, it makes use of full-scale avoid links and regular deep inspections that bring together fine-grained relevant information and general semantics mapped at multiple feature scales. A self-adaptive solid framework based on 2D and 3D U-Net forms the basis of the network. It utilises a trio of elementary U-Net models. The original U-Net is only significantly improved, and no extension plug-ins, such

as residual connections, dense interconnection, and various attention mechanisms, are used. In applications such as brain tumour segmentation, the nnU-Net produces unexpectedly accurate results [65].

Using prostate MRI volumes as input, our CNN is taught to predict segmentation for the entire volume simultaneously. To enhance training based on the Dice coefficient, they introduced a novel objective function. They add more data by using histogram matching and random non-linear transformations to make up for the small number of training sets. In order to classify Electrocardiogram (ECG) signals, [21] offered an alternative viewpoint based on Downsampling One-Dimensional-Local Binary Pattern (1D-DS-LBP) and LSTM. The approach takes into account the big-little relationship between adjacent nodes. Histograms of 1D local binary patterns are compiled and added to the LSTM model as per the anticipated approach. ECG signals, both unidirectional and bidirectional, are subjected to the LSTM. ECG signals from three different CHF states were used to evaluate the suggested methodology. As can be seen from the investigational results, the proposed method is able to achieve superior classification accuracies to those achieved by competing models.

From the discussions on earlier works, it is evident that there have been a few cases where DL has been used to diagnose structural heart disease using ECG [66]. As a result, cardiac standards can now be used to analyze the left endocardial ventricular border segments for wall motions, volume tests [67], chamber size analysis [68], valve mobility issues, and the appearance of pericardial effusion, among many other automated interpretations [69]. Each result attained a good level of performance. Unfortunately, not all research was shown on FC; only binary segmentation was performed. Due to the heterogeneity of specific lesion images, the diversity of heart anatomy between individuals, and the tiny target of diagnosing a defect of less than 2 *mm* in the cardiac muscle with low-pixel images, segmentation and fault diagnosis in the cardiac septal is challenging. These factors can lead to poor performance and numerous errors in segmentation and detection [70]. In addition, the earlier study on prenatal object detection was a medical case study [71]. Therefore, a thorough examination of CRD as a cardiac defect in fetuses is desirable. An attention network layer-based V-Net multitask learning model for FC image localization is proposed in this research; as such, a model can segment multi-regions, choose multi-patients, categorise multi-regions of interest, and detect complex health objects. In this work, segmentation, classification, and detection processes are simultaneously learned in order to make an accurate diagnosis of FCRD.

This research work is outlined as follows: The following section provides the details of data acquisition for this study: Section 3 presents the proposed methodology, followed by implementation and results from analysis in Section 4, and finally, this article is concluded in Section 5.

II. DATA ACQUISITION

In this analysis, two senior maternal-fetal doctors with expertise in FECG assessment at two general hospitals in Tamil Nadu, India, specified the key anatomic components utilised to assess picture quality. In the normal anatomy, the FC was viewed using all four standard views—4CH, 3VT, LVOT, and RVOT—whereas, in the aberrant anatomy, which includes disorders such as ASD, VSD, and AVSD, only the 4CH view was utilised to evaluate CRD. The FECG image was created using ultrasound video data from women who were between 18 and 24 weeks pregnant and had normal anatomy in the 4CH, 3VT, LVOT, and RVOT views. The GE Voluson E8 ultrasound scanner was used to capture the video in question. To identify the normal and pathological anatomical structures of the FC in utero, an study analysis with such a cross-sectional design has been done. The cross-sectional FC imaging for heart defect analysis exclusively uses the 4CH view since such a trial, as well as the ventricular portion of the FC, are clearly seen in this view.

For the purpose of calculating inter (or) intra-observer variability, a second annotator annotated a representative subset of videos. Since the heart's direction is uncertain, we created the contextual information by defining a circle with a radius of 80 pixels (the average value derived from metadata) all the way around the heart's manually marked centre. The view label linked to the frame was used to name the pixels inside the circle. Digital Imaging and Communications in Medicine (DICOM) formats were used to retrieve all FC images for retrospective analysis.

In this study, five conditions with the FCRD are investigated, including Fetal Rhabdomyomas noted in the Right Ventricle (FRRV) in approximately 134 images, Fetal Rhabdomyomas noted in the Left Ventricle (FRLV) in about 158 images, Fetal Rhabdomyomas noted in the Right Atrium (FRRA) in approximately 164 images, Fetal Rhabdomyomas noted in the Left Atrium (FRLA) in approximately 103 images, Fetal Rhabdomyomas noted in the Tricuspid Valve (FRTV) in approximately 78 images and around 273 images without any rhabdomyoma that means Normal Condition (NC). Out of the total 910 images, 80% of each class of images is used as a training set, and the remaining is used as a test set. All images were verified and annotated by two doctors to exclude human bias. Each image was documented explicitly by an old age person as well as the presence of a doctor with over five years of field experience. Again, they were followed by an associate chief doctor with expertise who evaluated and confirmed their findings. Each concerned individual provided written consent, and all data were customized to anonymize patient information.

A. DATA PREPARATION

Prior to analysis, there was the removal of coloured annotations on DICOM images, including profile traces, icons, texts, and callipers that contain patient Personal Health Information (PHI) (Fig. 1). On the picture borders, the identifying data

FIGURE 1. Identifying image annotations (IA) on a scan with Calipers, Icons, Text, and Profile Traces were among the IA.

FIGURE 2. Removal of IA on a scan with a CRD diagnosis.

were ridden by managing PHI using picture cropping. By first transforming the Red Green Blue (RGB) colour system that contains image data to a Hue Saturation Value (HSV) colour space, coloured annotations were eliminated. The H, S, and V values for the image pixels that belonged to a grey ultrasound image are empirically determined to vary from 0-27, 0-150, and 0-255. A coloured annotation, a component, is otherwise known as the pixels beyond specific ranges. There was a making of an image's Binary Mask (BM) in the third step with the labelling of annotation pixels as *"1"* and the ultrasound image pixels labelled *"0"*. The annotations' surrounding contours were then added to the BM using a 5 × 5 kernel dilation. Finally, the ultrasound image was artificially recreated without annotations by using the Navier-Stokes image infill approach (Fig. 2) [72].

After being cleaned, DICOM images are transformed into grayscale (1 channel) images. To ensure more stability during Neural Network (NN) training, normalization of intensities was done to have a close to '0' mean and one standard deviation [73]. The photos were finally downsized to 256 x 256 pixels.

All annotations on the images have been erased as a result of the Navier-Stokes image infill process.

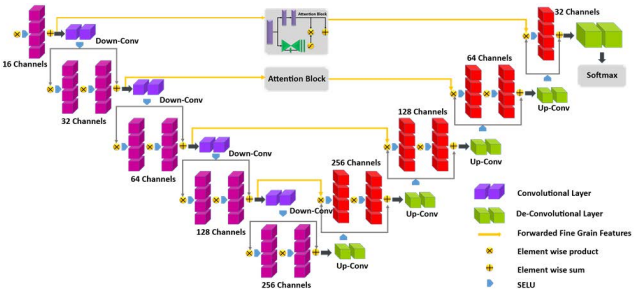

FIGURE 3. ARVNet architecture.

III. METHODOLOGY

A. OUR PROPOSED ARCHITECTURE'S OUTLINE

We built our ARVNet by stacking several attention networks (Fig. 3). The Mask Branch (MB) as well as Trunk Branch (TB), are the two divisions that are made for each Attention Network. The TB handles feature processing and is adaptable to all modern network architectures. In this study, we build the attention network using the pre-activation residual unit [74], ResNeXt [75], and Inception [76] as our ARN basic units. The MB learns the same size mask M(x), which soft-weights the output features "T" from the TB output "T(x)" to input "x" [77]. The quick feedforward process, as well as the feedback attention process, are modelled by the bottom-up, top-down organizational structure.

1) ATTENTION RESIDUAL NETWORK (RESNET)

According to [78], an output mask is being used to control gates for the neurons of the TB of the Highway Network. Attention Network H's output is EQU (1):

$$H_{i,c}(x) = M_{i,c}(x) * T_{i,c}(x) \tag{1}$$

where all discrete points are represented by 'i' and the channel index is $c \in \{1, .., C\}$. The entire structure can make use of end-to-end training.

A basic approach would create a soft weight mask using a single network branch, similar to such a spatial transformer layer, rather than adding an attention network (Fig. 4) to our architecture.

(i) An exponential number of channels would have been needed to record all groupings of distinct elements in a single MB.

(ii) Each attention network only makes one change to its characteristics.

In the event that any sections of the image cannot be modified, the following network modules would not be given another chance. These issues are resolved by an ARN. In the Attention Network, each TB would have a separate MB to teach attention that is tailored to its features.

2) VNET ARCHITECTURE

The suggested V-Net method is split into two segments: a path for compression on the left and a path for decompression on the right that decelerates its input until it returns to its original

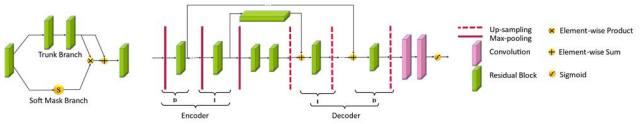

FIGURE 4. Attention ResNet.

size. Every one of the convolutions is being carried out with the proper padding, with the dual objectives of exploiting input qualities first and lowering resolution by using the appropriate stride at the end of each stage. There are some differences between the recommended network design for V-Net and the widely used U-Net design. There are various phases that function at different resolutions that make up the VNet architecture, mainly on the left. One, two, and three convolutional layers are present at each level by structuring each step in a similar way to the one described. We are capable of studying the residual function within all phases. For the need to train the residual function, data input for each stage is being treated using non-linearities present inside CL; besides, these non-linearities are attached to the last CL for the output phase. Such network guarantees for convergence are mainly compared to non-residual-trained designs like U-Net. Each level also uses Convolutional Layers (CL) with just a 5×5 grid size.

The convolution process is depicted in EQU (2).

$$x_4(ii, ij) == \sum_{n=1}^{N} \left\{ \sum_{p=0}^{w_{j-1}h_{j-1}} \sum_{q=0} x_n \times (i.s_j + p, j.s_j + q).h(p, q) \right\} + b_k \quad (2)$$

$x_k(ii, ij)$ is the representation of the pixel rate (ii, ij) in the k^{th} filter size for the input map; furthermore, (ii, ij) and $x_n(ii, ij)$ are the pixel rate in the network, whereas for k^{th} filter that is exchanged between every location (p, q), b_k will be a bias parameter; s_f will be the *sampling stride,* $h_k(p, q)$. Data resolution is minimised as it passes through subsequent steps outside of the compression path, and this is done with the aid of such a CL with size 22 and strides 2. The sizes for such resulting feature maps are decreased to be a result of a 2^{nd} operation's evaluation of 22 non-overlapping patches on its own and the subsequent extraction of features. To obtain the same result as the pooling layers suggested by [79], max-pooling layers are substituted in our method through CL. We applied such convolutional techniques to double the amount of the feature maps.

Because of the residual framework construction of the approach and the fact that at each phase, the feature channel numbers double at V-Net compression paths. The network's memory footprint exists while training is minimised by using CL instead of layer pooling. Although we tried to save as many of the features as possible, they were down-sampled even during the pooling step. CL have several advantages over pooling layers in the suggested method, including the ability to process input at a more excellent resolution while also detecting minute details and gathering additional context data through broadening the viewpoint of the input data [80]. Numerous functions, including tanh and corrected functions, can be used for the convolution procedure (EQU (3)). The downsampling stage expands a receptive field for these features that are examined in the network's subsequent layers while decreasing the amount of input. Inside the left half of a network, a sum of features examined through each stage are two times bigger than the one inside the layer before it.

$$Z(x_k(ii, jj)) = f(\sum_{k=1}^{k} x_k(ii, jj).w_k + b_k) \Leftrightarrow Z = f(X.W + b) \quad (3)$$

The result of a convolution operation is $x_k(ii, jj)$, which is input to a NN's AF. Here 'w' appears to be the weight vector, and 'b' to be, in fact, the bias vector. We make use of *'Scaled Exponential Linear Units (SELUs)'*, which make a NN self-normalizing for such AF and are offered in [81]. The following is a definition of the SELU-AF:

$$SELU(x) = \gamma\{^{x}_{\alpha e^x} | IF^{x \geq 0}_{x \leq 0}\} \quad (4)$$

where alpha, as well as gamma, seem to be constant parameters that are learned from input and remain the same throughout the network's iterations. For normally scaled inputs, the values are 1.6732 and 1.0507, respectively. Building a g mapping using attributes leading towards SNNs is made easy by SELUs. SNNs cannot be calculated using leaky Rectified Linear Units (ReLUs), sigmoid units, tanh units, or (scaled) ReLUs. The Activation Function (AF) needs to have the following limits: (a) *'-ve'* and *'+ve'* values to control the average; (b) overload regions (derivatives closer to '0') to reduce change if it is enormous in the end layer; and (c) a slope more significant than that to increase variance if it is tiny in the end layer. Whenever variance damping is balanced by raising the variance, the latter assures a fixed point. We increased the same Exponential Linear Unit (ELU) [82] $\gamma > 1$ to ensure that the slope of such an AF is larger than that of *'+ve'* net inputs.

3) LOSS FUNCTION

By allocating each voxel 'x' with an AF value $p(x) \in [0, 1]$p(x) $\in [0, 1]$, predicting each voxel's class is the main goal with respect to medical image volume. For learning a prediction model $\phi(x; \theta) : x \rightarrow p(x)$, a DL technique is adopted, where the model parameters are denoted by 'θ' and, for organ/class 'i', p_i is the activation value.

4) LOSS FUNCTION OF CROSS-ENTROPY

A computation of CE loss could be done $C = \sum_x \sum_i t_i l_n(p_i(x))$ for multi-class issues; the forecasted Probability Mass Function (PMF) that allocates each class for each voxel with a probability/AF value is represented by 'p'. Neither the one-hot encoding target nor ground truth PMF is represented by 't', where over the number of organs, the index 'i' iterates as well as 'x' iterates upon many samples or voxels. The computation of the total of various binary

cross entropy terms is the computation of 'C'. There is feasibility in controlling (FP)/(FN) for certain multi-class issues similar to this paper. The rewriting of 'C' can be $\sum_x t_i l_n(p_i) + (1 + t_i) l_n(1 - p_i)$ in the case of binary classification. There is an accurate prediction since FP is penalized by the term $(1 + t_i) l_n(1 - p_i)$ '0'. The multi-class issues such as $\frac{1}{N} \sum_{i=1}^{N} t_i l_n(p_i) + (1 + t_i) l_n(1 - p_i)$ can make use of the binary formulation that is also capable of being extended.

Where, $N = \text{Classes} \times TestSamples$.

So, the multiple binary cross entropies' average is nothing but the output.

5) DICE OPTIMIZATION FUNCTION

The image separation precision that could be composed in the types of $Dice = TP/\text{Number of Positive} + \text{Number of NP}$ or $Dice = 2xTP/(FN + 2xTP) + FP$ is evaluated by using the Dice function, a predominantly used metric. In order to generalize into multi-class issues, that shall similarly be rewritten as the weighted function [83]. But in the above formulations, it is impossible to control the fining of neither FPs nor FNs and their trade-off. In the following way, a generalized/weighted DBLF in the binary case is defined as:

$$GDL = 1 \frac{2(\sum_{l=1}^{2} wl \sum_n r_{\ln} p_{\ln})}{\sum_{l=1}^{2} wl \sum_n r_{\ln} p_{\ln}} \tag{5}$$

where voxel values r_n in the reference foreground segmentation and voxel values p_n in forecasted segmentation are represented by 'R' and 'P'.

6) HYBRID LOSS FUNCTION

We learn the weights by lowering the loss function. It is utilised to derive a weighted sum among dice losses and an altered CE from such a combo loss function [84]. Our loss 'L' is set up to efficiently use the Dice function, which manages the issue of input class imbalance, such that the classification of a tiny frame is done from a vast backdrop. This is similar to a weighted sum of two relations. Together, it normalizes the trade-off between FP and FN and applies light training using CE, as was stated previously. The following is how EQU (6) represents it as a dice loss:

$$L = \alpha(-\frac{1}{N} \sum_{i=1}^{N} \beta(t_i l_n p_i) + (1 - \beta)[(1 - t_i) l_n(1 - p_i)] - (1 - \alpha) \sum_{i=1}^{K} \left[\frac{2 \sum_{i=1}^{N} p_i t_i + s}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} t_i + s} \right] \tag{6}$$

when 'β' is less than 0.5, the penalty for FP is greater than the penalty for FN in the loss function 'L', where it limits the flow for Dice term contribution and model penalization for FP or FN regulation $\beta \in [0, 1]$ due to the apparent term $(1 - t_i)$. Inversely, weight $(1 - t_i)$ is heavier.

According to [43], add-one smoothing, which involves adding the unity constant S for both a denominator as well as a numerator of a dice term, is used to prevent division by *'0'*. Add-one smoothening is the subset of the Additive/Laplace/Lidstone smoothing. Despite the fact that the proposed loss is just the simple combination of two distinct loss functions, a binary form of the CE was chosen to represent the apparent strengthening of a conscious trade-off between FP and negatives. The dice term, which provided the global spatial data, is also employed to keep the model parameters from being affected by harmful local minima. After sigmoid normalisation over these channels, like classes in the last layer, the flat quantities (for both forecasted as well as ground truth volumes containing different objects, one-hot multi-label encoding is carried out) are used to calculate a hybrid loss function of WHDC size, in which the height, depth, and quantities of channel classes seem to be W, H, D, and C, respectively. Direct control of FPs, as well as FNs across the entire volume, is possible in this method, which is easily adapted to multi-class segmentation.

TABLE 1. ARVNet architecture.

ENCODER	OUTPUT SIZE	DECODER	PRE-OPERATION	OUTPUT SIZE
Input	224 x 224 x 32 x 1	Att1	[Res4], depth=0	14 x 14 x 2 x 256
Conv1	224 x 224 x 32 x 32	Res7	[Up1, Att1]	14 x 14 x 2 x 256
Pooling	112 x 112 x 16 x 32	Up2		28 x 28 x 4 x 256
Rest	112 x 112 x 16 x 32	Att2	[Res3], depth=1	28 x 28 x 4 x 128
Pooling	56 x 56 x 8 x 32	Res8	[Up2,Att2]	28 x 28 x 4 x 128
Res2	56 x 56 x 8 x 64	Up3		56 x 56 x 8 x 128
Pooling	56 x 56 x 4 x 64	Att3	[Res2], depth=2	56 x 56 x 8 x 64
Res3	28 x 28 x 4 x 128	Res9	[Up3,Att3]	56 x 56 x 8 x 64
Pooling	14 x 14 x 2 x 128	Up4		112 x 112 x 16 x 64
Res4	14 x 14 x 2 x 256	Att4	[Res1], depth=3	112 x 112 x 16 x 32
Pooling	7 x 7 x 1 x 256	Res10	[Up4, Att4]	112 x 112 x 16 x 32
Res5	7 x 7 x 1 x 512	Up5		224 x 224 x 32 x 32
Res6	7 x 7 x 1 x 512	Conv2	[Up5, Conv1]	224 x 224 x 32 x 32
Up1	14x14x2x512	Conv3		224 x 224 x 32 x 1

B. UTILIZATION OF 2D-ARVNET FOR CRD LOCALIZATION

For segmenting the area of interest, a new 2D-ARVNet is introduced, which can remove redundant data, reduce the computational cost of the subsequent 3D-ARVNet, and provide more helpful information. Finding the boundary box was the first stage's fundamental goal. It is used as a *"baseline"* to condense the areas of interest for CRD localization. We downscaled the chunks to 256 × 256, preprocessed them, and then input those into such a trained 2D-ARVNet model. The pieces are then stacked in the order of their actual status. Then, according to [44], 3D connected-component labelling was applied. In order to identify specific areas and gauge the size of regions, each connected component in an image is given a unique label through this technique. The largest component was then determined to be the coarse CRD area. To conclude, we intercalated the CRD region back to its previous volume size using an in-plane resolution of

512 × 512. Connected module labelling is the process of allocating a unique label for every connected module inside an image.

C. SEGMENTATION USING 3D-ARVNET

A 3D-ARVNet was actually a 3-D model which fully incorporated spatial and volume information. A 3D-ARVNet-type framework [85] blends low-resolution as well as high-resolution data for creating precise segmentation (Tab. 1).

In contrast, training with large picture patches provides significantly more contextual data than training with smaller patches, which typically yields superior overall segmentation results. Following is an explanation of the segmentation process:

The 3D-ARVNet contains fewer parameters than the traditional U-Net. In this architecture, a set of the mentioned parameters have already been decreased, primarily the 4M training parameters. A 3D-ARVNet has been used to create 3D CRD probability patches sequentially on each patch of a US image. When interpolating a CRD bounding box to a fixed size, such as 224 × 224, we created the training patches in the (x, y, z) planes. 32 chunks were randomly chosen in the 'z' direction. A boundary box's original size was then recreated by interpolating and stacking the probability patches. A voting technique from overlapped sub-patches was used to determine a ROI's final CRD probability. The 3D connected-component labelling has been used to produce the final CRD region, and the most significant component on the combined ROI was selected.

IV. MODEL TRAINING AND IMPLEMENTATION DETAILS

The TensorFlow [86] and Keras [87] libraries were used to build the ARVNet architecture. All the models had initial training. The network's parameters were given random initial values before being trained using Adam-based back-propagation [88] with initial Learning Rates (LR) of 0.001, $\beta 1 = 0.9$, and $\beta 2 = 0.999$. If the network plateaued after 20 epochs, the LR would be decreased to LRx0.1. On the training dataset, we applied 5-fold cross-training and assessed the results on the test dataset [89], [90], [91]. The total number of epochs for the heart image, as well as FCRD training, are set at 50 and 50 at each fold, respectively. All prediction results from the five models are ensembled using an integration process based on voting. An NVIDIA 1080Ti GPU was used to train all the models. In our trials, training an epoch of our 3D ARVNet to FCRD segmentation takes approximately 100/40 minutes, respectively.

A. EXPERIMENTAL RESULTS

Accuracy, Specificity, Precision, Recall, and F1-score are used to assess performance and identify ASDs correctly. Dice Similarity Coefficient (DSC) is used as the evaluation process metric during cardiac anatomical segmentation. These are their definitions, EQU (7), EQU (8), EQU (9), EQU (10), EQU (11), and EQU (12):

$$Accutracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{7}$$

$$\text{Recall} = Sensitivity = \frac{TP}{TP + FN} \tag{8}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{9}$$

$$Specificity = \frac{TN}{TN + FP} \tag{10}$$

$$F1 - score = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision+Recall}} \tag{11}$$

$$DSC = \frac{2 \times |A \cap B|}{|A| + |B|} \tag{12}$$

The counts of True Positives, False Positives, True Negatives, and False Negatives, respectively, are TP, FP, TN, and FN among them. TP and TN stand for the advantages and disadvantages of accurate predictions in relation to the actual data. FP and FN stand for the positives and negatives, respectively, of inaccurate forecasts in relation to the actual data. The F1-score, which has a value between 0 and 1, seems to be the harmonic average of Precision and Recall. The model performs better with a higher value. The segmented area is termed "B", while the ground truth area is termed "A".

TABLE 2. Performance of the proposed ARVNet model.

Method	Sp	DSC	Acc	Precision	Recall	FI
V-Net	0.9962	0.9969	0.9957	0.9663	0.9485	0.9729
CNN	0.8341	0.9978	0.9252	0.9908	0.9444	0.9613
FCN	0.9902	0.9914	0.9335	0.8971	0.9163	0.9066
Mask-RCNN	0.9913	0.9924	0.9365	0.9788	0.8413	0.9099
ARVNet **Proposed**	0.9972	0.9984	0.9959	0.9985	0.9726	0.9876

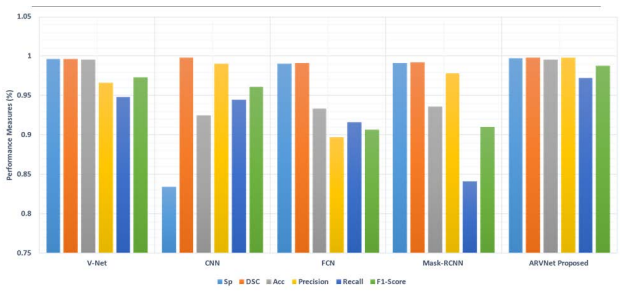

FIGURE 5. Comparison of the proposed vs other models as FCRD segmentation.

B. COMPARATIVE ANALYSIS OF DIFFERENT NETWORK ARCHITECTURES

The proposed method is compared against four cutting-edge approaches to multitask learning. Tab. 2 displays the output from the CNN-based ASD detection model [92], FCN-based CHD detection proposed by [93], a Mask-RCNN-based DL model for handling fetal ultrasonography images for

diagnosing ASD by [94], and a method for the segmentation of 2D-CT images of fetal skulls based on a V-Net network model by [95].

We get average values of 0.99 for specificity and dice using ARVNet. The normal values for the F1-score, recall, and accuracy is 0.98, 0.97, and 0.99, respectively. Fig. 5 and Table 1 show a suggested framework that performs better than most advanced NNs. Modern NNs and the ARVNet approach were compared using a one-way ANOVA, which revealed statistical differences in mean Specificity as well as Dice (p=0.32). Fig. 6 (a) ∈, presents error (%) in detections for comparisons between the proposed model and the other models for FCRD predictions over five classes, including FRTV, FRLA, FRRA, FRLV and FRRV.

C. BENCHMARK ARVNet

We benchmarked our recommended ARVNet with three state-of-the-art architectures: Focusnet++, ENNet, and nnU-Net. All networks were trained using our training medical data to generate the probability measures for each class label. We test the fetal heart using 4 standard views—4CH, 3VT, LVOT, and RVOT—whereas, in the aberrant anatomy, which includes disorders such as ASD, VSD, and AVSD, only the 4CH view was utilised to evaluate CRD. The FECG image was created using ultrasound video data from women who were between 18 and 24 weeks pregnant and had normal anatomy in the 4CH, 3VT, LVOT, and RVOT views. The outcome of the proposed ARVNet against the other models in FCRD segmentation is referred to in Tab. 3, which shows that the ARVnet outperformed Focusnet++, ENNet and nnU-Net in terms of the Sp, DSC, Acc, Precision, Recall and F1-score. ARVNet produced average Precision, Acc, and Dsc values of 99.7%, 99.8%, 99.5%, 99.8%, 97.2%, and 98.7% respectively. All results exceeded 90%, given that the baseline for all metrics was 90%. Therefore, the ARVnet model could detect all classes in the four views.

D. GRADIENT-WEIGHTED CLASS ACTIVATION MAPPING (GRAD-WCAM)

In order to enhance the interpreting skills in our trained ARVNet model, the visual contextualizes significant elements inside the image data, which were utilized for predictions by the model; this was done using GRAD-WCAM [96]. Grad-WCAM is a method that is frequently used to illustrate DL algorithms. The areas of every image, which were crucial for the prediction of the model, were highlighted by creating heat maps using 8 × 8 feature maps (Fig. 7).

The FCRD were primarily highlighted in the GRAD-WCAM heat maps (Fig. 8). Certain heat maps indicated Rhabdomyoma at the tricuspid valve annulus protruding across the tricuspid valve (Fig. 9) and heat map indicates FCRD within the pericardial space and compressing the right ventricle, there were instances of inadequate localisation.

FCRD instances are in the second row, while normal NT cases are in the top row. Regions of great importance are highlighted in orange-red, whereas regions of lower or no

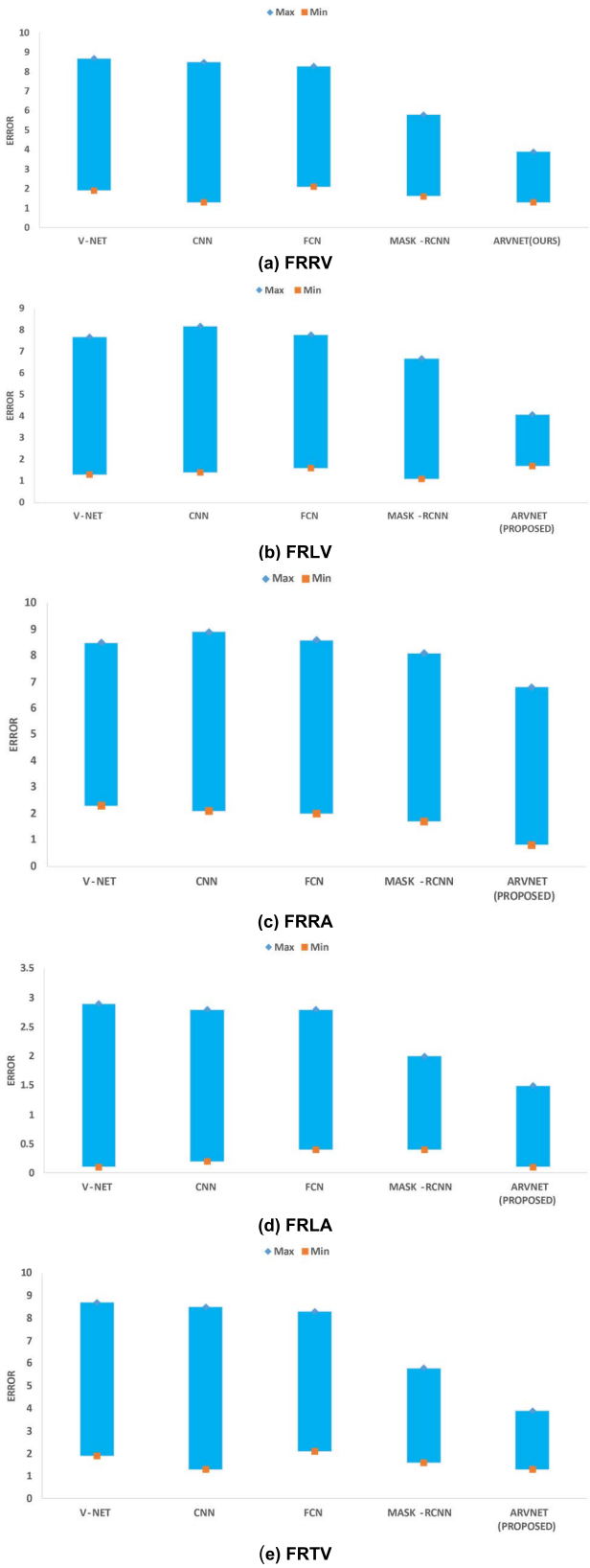

FIGURE 6. (a)-(e): Comparison of detection error for the proposed model against other models.

importance are highlighted in blue. For the CRD images, the G-CAM heatmaps, which emphasize the area around the

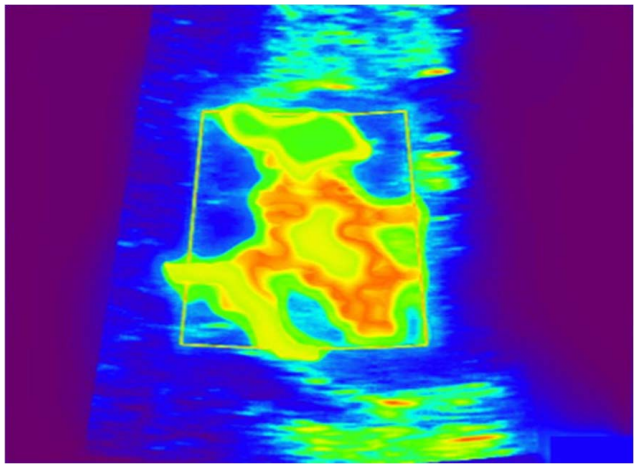

FIGURE 7. Grad-CAM image of a FCRD case. The image's area which had the most impact on the model's prediction is highlighted in orange-red.

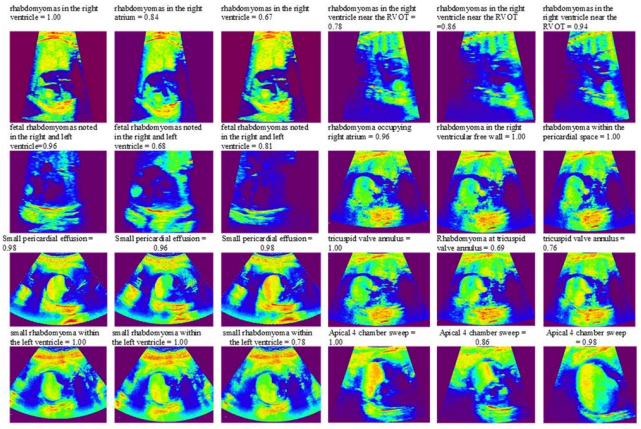

FIGURE 8. Grad-CAM heat maps for some sample images from each class of test set.

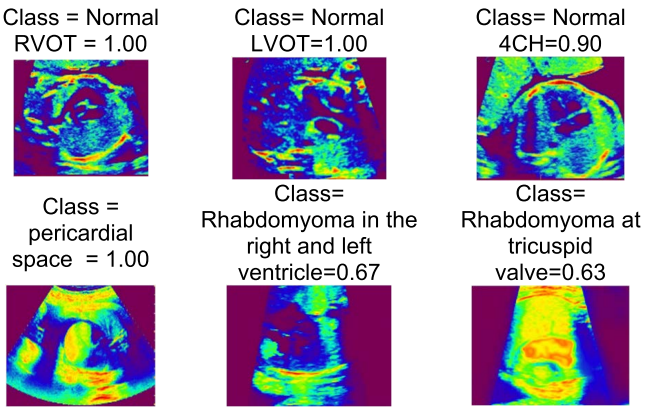

FIGURE 9. Exemplary GRAD-CAM heat map.

tumor, are incorporated into the ideal model. (a & b) In the normal NT situation that possesses good localization (TN), the class with a high (1.00) probability of output is correctly identified by the model. (c) Normal NT case with weak localization, with a probability of output of 0.90 (d) A case of a CRD in pericardial space with excellent localization, where the class with a high (1.00) probability of output (TP) was correctly identified by the model. (C) A case of a CRD with weak localization, with a probability of output of 0.67, where the model identified this class wrongly (FP). (D) Example of a CRD with poor localisation, where the model correctly identified the class, but the probability of output refers to uncertainty (0.63) (TP).

TABLE 3. Benchmarking the proposed ARVNet model.

Architecture	View	Sp	DSC	Acc	Precision	Recall	F1-Score
Focusnet++	4CH	0.9964	0.9979	0.9874	0.98	0.9725	0.9678
	3VT	0.994	0.9955	0.985	0.9776	0.9701	0.9654
	LVOT	0.9935	0.995	0.9845	0.9771	0.9696	0.9649
	RVOT	0.9969	0.9984	0.9879	0.9805	0.973	0.9683
ENNet	4CH	0.9979	0.9983	0.9885	0.9813	0.9723	0.9704
	3VT	0.9955	0.9959	0.9861	0.9789	0.9699	0.968
	LVOT	0.995	0.9954	0.9856	0.9784	0.9694	0.9675
	RVOT	0.9984	0.9988	0.989	0.9818	0.9728	0.9709
nnU-Net	4CH	0.9964	0.9981	0.9735	0.9808	0.9611	0.9726
	3VT	0.994	0.9957	0.9711	0.9784	0.9587	0.9702
	LVOT	0.9935	0.9952	0.9706	0.9779	0.9582	0.9697
	RVOT	0.9969	0.9986	0.974	0.9813	0.9616	0.9731
ARVNet	4CH	0.9984	0.9996	0.9971	0.9997	0.9738	0.9888
	3VT	0.996	0.9972	0.9947	0.9973	0.9714	0.9864
	LVOT	0.9955	0.9967	0.9942	0.9968	0.9709	0.9859
	RVOT	0.9989	0.9999	0.9976	1.0002	0.9743	0.9893

V. CONCLUSION AND FUTURE WORK

Our primary goal of developing an algorithm to segment FCRD from ultrasound images automatically was successfully accomplished. Because it is modular, there are better chances for quality assessment, and it is possible to make a significant improvement on several aspects at once. As far as we are familiar with the idea, this is the first report to detail the extraction of FCRD from an ultrasound image. In this research study, we show that FCRD segmentation can be taken into account by a V-Net architecture that is based on ARV.

The proposed V-Net architecture has been shown to segment both normal and FCRD hearts successfully. High precision and recall values are also present, demonstrating that images are not over- or under-segmented. The outcomes for the game of dice were similarly impressive. MIS using the recommended reference architecture is very close to ground truth when the dice score is high. Our study's performance results are better than those of other models' research, but there has been no other research project like it.

The proposed FCRD detection system was developed using a training dataset of 820 images belonging to 6 classes, such as (a) Fetal Rhabdomyomas noted in the Right Ventricle (FRRV), (b) Fetal Rhabdomyomas noted in the Left Ventricle (FRLV), (c) Fetal Rhabdomyomas noted in the Right Atrium (FRRA), (d) Fetal Rhabdomyomas noted

in the Left Atrium (FRLA), (e) Fetal Rhabdomyomas noted in the Tricuspid Valve (FRTV) and (f) images without any rhabdomyoma that means Normal Condition (NC). The proposed system was able to accurately identify FCRD from FECD images, as demonstrated by independent test set experiments; this paves the way for future AI-based diagnostic testing of FCRD. In this investigation, we talked about how the proposed ARVNet was able to tell the difference between normal heart classes and CRD classes with a 99.59% level of accuracy.

It would be ideal for conducting more ablation studies on the influence of the model on final FCRD detection. However, due to the high computational cost of data labelling, our current independent test data only has FCRD labels for each image without segmentation ground truth. As such, we recognize this as a shortcoming and aim to address it in our future work, in addition to extending the scope of this investigation to include other abnormal conditions that could potentially contribute significantly to this field of study. In addition, the defect size parameter was not considered in this research investigation; for future research, it will be crucial to diagnose the condition's serious nature.

ACKNOWLEDGMENT

The authors would like to thank Selvam Hospital, Melapalayam, Tirunelveli, Tamil Nadu, India, was provided medical datasets for conducting their research work as successful manner.

REFERENCES

[1] L. Chen, Y. Jiang, and J. Wang, "Fetal cardiac rhabdomyoma due to paternal mosaicism of TSC2: A case report," *Medicine*, vol. 99, no. 35, Aug. 2020, Art. no. e21949.

[2] P. Sciacca, V. Giacchi, C. Mattia, F. Greco, P. Smilari, P. Betta, and G. Distefano, "Rhabdomyomas and tuberous sclerosis complex: Our experience in 33 cases," *BMC Cardiovascular Disorders*, vol. 14, no. 1, pp. 1–11, Dec. 2014.

[3] T. Ide, T. Miyoshi, S. Katsuragi, R. Neki, K.-I. Kurosaki, I. Shiraishi, J. Yoshimatsu, and T. Ikeda, "Prediction of postnatal arrhythmia in fetuses with cardiac rhabdomyoma," *J. Maternal-Fetal Neonatal Med.*, vol. 32, no. 15, pp. 2463–2468, Aug. 2019.

[4] P. B. Crino, K. L. Nathanson, and E. P. Henske, "The tuberous sclerosis complex," *New England J. Med.*, vol. 355, pp. 56–1345, Sep. 2006.

[5] P. Curatolo, R. Bombardieri, and S. Jozwiak, "Tuberous sclerosis," *Lancet*, vol. 372, pp. 68–657, Jan. 2008.

[6] A. S. Chao, A. Chao, T. H. Wang, Y. C. Chang, Y. L. Chang, C. C. Hsieh, R. Lien, and W. J. Su, "Outcome of antenatally diagnosed cardiac rhabdomyoma: Case series and a meta-analysis," *Ultrasound Obstetrics Gynecol.*, vol. 31, no. 3, pp. 289–295, 2008.

[7] E. Ekmekci, B. O. Ozkan, M. S. Yildiz, and B. Kocakaya, "Prenatal diagnosis of fetal cardiac rhabdomyoma associated with tuberous sclerosis: A case report," *Case Rep. Women's Health*, vol. 19, Jul. 2018, Art. no. e00070.

[8] L. Wu, B. Dong, X. Liu, W. Hong, L. Chen, K. Gao, Q. Sheng, Y. Yu, L. Zhao, and Y. Zhang, "Standard echocardiographic view recognition in diagnosis of congenital heart defects in children using deep learning based on knowledge distillation," *Frontiers Pediatrics*, vol. 9, Jan. 2022, Art. no. 770182.

[9] W. Hong, Q. Sheng, B. Dong, L. Wu, L. Chen, L. Zhao, Y. Liu, J. Zhu, Y. Liu, Y. Xie, Y. Yu, H. Wang, J. Yuan, T. Ge, L. Zhao, X. Liu, and Y. Zhang, "Automatic detection of secundum atrial septal defect in children based on color Doppler echocardiographic images using convolutional neural networks," *Frontiers Cardiovascular Med.*, vol. 9, Apr. 2022, Art. no. 834285.

[10] J. Zhang, S. Gajjala, P. Agrawal, G. H. Tison, L. A. Hallock, L. Beussink-Nelson, M. H. Lassen, E. Fan, M. A. Aras, C. Jordan, K. E. Fleischmann, M. Melisko, A. Qasim, S. J. Shah, R. Bajcsy, and R. C. Deo, "Fully automated echocardiogram interpretation in clinical practice," *Circulation*, vol. 138, no. 16, pp. 1623–1635, Oct. 2018.

[11] Z. Sobhaninia, S. Rafiei, A. Emami, N. Karimi, K. Najarian, S. Samavi, and S. M. R. Soroushmehr, "Fetal ultrasound image segmentation for measuring biometric parameters using multi-task deep learning," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Berlin, Germany, Jul. 2019, pp. 6545–6548.

[12] R. Vullings, "Fetal electrocardiography and deep learning for prenatal detection of congenital heart disease," in *Proc. Comput. Cardiology Conf. (CinC)*, Singapore, Dec. 2019, p. 1.

[13] K. Vo, T. Le, A. M. Rahmani, N. Dutt, and H. Cao, "An efficient and robust deep learning method with 1-D octave convolution to extract fetal electrocardiogram," *Sensors*, vol. 20, no. 13, p. 3757, Jul. 2020.

[14] D. Ravi, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G. Z. Yang, "Deep learning for health informatics," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 1, pp. 4–21, Dec. 2017.

[15] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3462–3471.

[16] R. Arnaout, L. Curran, Y. Zhao, J. Levine, E. Chinn, and A. Moon-Grady, "Expert-level prenatal detection of complex congenital heart disease from screening ultrasound using deep learning," medRXiv, 2020, doi: 10.1101/2020.06.22.20137786.

[17] J. Torrents-Barrena, G. Piella, N. Masoller, E. Gratacós, E. Eixarch, M. Ceresa, and M. Á. G. Ballester, "Segmentation and classification in MRI and U.S. fetal imaging: Recent trends and future prospects," *Med. Image Anal.*, vol. 51, pp. 61–88, Jan. 2019.

[18] H. C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, Feb. 2016.

[19] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention* (Lecture Notes in Computer Science), vol. 9351, N. Navab, J. Hornegger, W. Wells, and A. Frangi, Eds. Cham, Switzerland: Springer, 2015.

[20] O. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Medical Image Computing and Computer-Assisted Intervention* (Lecture Notes in Computer Science), vol. 9901, S. Ourselin, L. Joskowicz, M. Sabuncu, G. Unal, and W. Wells, Eds. Cham, Switzerland: Springer, 2016.

[21] 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)*, Stanford, CA, USA, Oct. 2016, pp. 565–571, doi: 10.1109/3DV.2016.79.

[22] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[23] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. DLMIA ML-CDS 2017* (Lecture Notes in Computer Science), vol. 10553. Cham, Switzerland: Springer, 2017, doi: 10.1007/978-3-319-67558-9_28.

[24] M. Fatima and M. Pasha, "Survey of machine learning algorithms for disease diagnostic," *J. Intell. Learn. Syst. Appl.*, vol. 9, no. 1, pp. 1–16, 2017, doi: 10.4236/jilsa.2017.91001.

[25] M. Kim, J. Yun, Y. Cho, K. Shin, R. Jang, H.-J. Bae, and N. Kim, "Deep learning in medical imaging," *Neurospine*, vol. 16, no. 4, pp. 657–668, 2019.

[26] N. Shahid, T. Rappon, and W. Berta, "Applications of artificial neural networks in health care organizational decision-making: A scoping review," *PLoS ONE*, vol. 14, no. 2, Feb. 2019, Art. no. e0212356.

[27] I. Castiglioni, L. Rundo, M. Codari, G. Di Leo, C. Salvatore, M. Interlenghi, F. Gallivanone, A. Cozzi, N. C. D'Amico, and F. Sardanelli, "AI applications to medical images: From machine learning to deep learning," *Phys. Medica*, vol. 83, pp. 9–24, Mar. 2021.

[28] M. Alzubaidi, H. D. Zubaydi, A. A. Bin-Salem, A. A. Abd-Alrazaq, A. Ahmed, and M. Househ, "Role of deep learning in early detection of COVID-19: Scoping review," *Comput. Methods Programs Biomed.*, vol. 1, Jan. 2021, Art. no. 100025.
[29] M. S. Alzubaidi, U. Shah, H. D. Zubaydi, K. Dolaat, A. A. Abd-Alrazaq, A. Ahmed, and M. Househ, "The role of neural network for the detection of Parkinson's disease: A scoping review," *Healthcare*, vol. 9, no. 6, p. 740, Jun. 2021.
[30] H. Fujita, "AI-based computer-aided diagnosis (AI-CAD): The latest review to read first," *Radiol. Phys. Technol.*, vol. 13, no. 1, pp. 6–19, Mar. 2020.
[31] X. Tang, "The role of artificial intelligence in medical imaging research," *BJR Open*, vol. 2, no. 1, Nov. 2020, Art. no. 20190031.
[32] G. A. Cheikh, A. B. Mbacke, and S. Ndiaye, "Deep learning in medical imaging survey," in *Proc. CEUR Workshop*, vol. 2647, 2020, pp. 111–127.
[33] R. C. Deo, "Machine learning in medicine," *Circulation*, vol. 132, pp. 1920–1930, Nov. 2015.
[34] G. S. Handelman, H. K. Kok, R. V. Chandra, A. H. Razavi, M. J. Lee, and H. Asadi, "EDoctor: Machine learning and the future of medicine," *J. Internal Med.*, vol. 284, no. 6, pp. 603–619, Dec. 2018.
[35] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: Review, opportunities and challenges," *Briefings Bioinf.*, vol. 19, no. 6, pp. 1236–1246, 2017.
[36] J. Balayla and G. Shrem, "Use of artificial intelligence (AI) in the interpretation of intrapartum fetal heart rate (FHR) tracings: A systematic review and meta-analysis," *Arch. Gynecol. Obstetrics*, vol. 300, no. 1, pp. 7–14, Jul. 2019.
[37] Z. Chen, Z. Liu, M. Du, and Z. Wang, "Artificial intelligence in obstetric ultrasound: An update and future applications," *Frontiers Med.*, vol. 8, p. 1431, Aug. 2021.
[38] P. Iftikhar, M. V. Kuijpers, A. Khayyat, A. Iftikhar, and M. D. De Sa, "Artificial intelligence: A new paradigm in obstetrics and gynecology research and clinical practice," *Cureus*, vol. 12, no. 2, p. e7124, Feb. 2020, doi: 10.7759/cureus.7124.
[39] R. Arnaout, L. Curran, Y. Zhao, J. C. Levine, E. Chinn, and A. J. Moon-Grady, "An ensemble of neural networks provides expert-level prenatal detection of complex congenital heart disease," *Nature Med.*, vol. 27, no. 5, pp. 882–891, May 2021.
[40] B. Raef and R. Ferdousi, "A review of machine learning approaches in assisted reproductive technologies," *Acta Inf. Med.*, vol. 27, pp. 205–211, 2019.
[41] P. J. Correa, Y. Palmeiro, M. J. Soto, C. Ugarte, and S. E. Illanes, "Etiopathogenesis, prediction, and prevention of preeclampsia," *Hypertension Pregnancy*, vol. 35, no. 3, pp. 280–294, Jul. 2016.
[42] E. C. Larsen, O. B. Christiansen, A. M. Kolte, and N. Macklon, "New insights into mechanisms behind miscarriage," *BMC Med.*, vol. 11, no. 1, pp. 1–10, Dec. 2013.
[43] C. Sen, "Preterm labor and preterm birth," *J. Perinatal Med.*, vol. 45, no. 8, pp. 911–913, Jan. 2017.
[44] Z. Chen, Z. Wang, M. Du, and Z. Liu, "Artificial intelligence in the assessment of female reproductive function using ultrasound: A review," *J. Ultrasound Med.*, vol. 41, no. 6, pp. 1343–1353, Jun. 2022.
[45] D. Avola, L. Cinque, A. Fagioli, G. Foresti, and A. Mecca, "Ultrasound medical imaging techniques: A survey," *ACM Comput. Surv.*, vol. 54, no. 3, p. 38, Apr. 2022, Art. no. 67, doi: 10.1145/3447243.
[46] S. Masood, M. Sharif, A. Masood, M. Yasmin, and M. Raza, "A survey on medical image segmentation," *Current Med. Imag. Rev.*, vol. 11, no. 1, pp. 3–14, Apr. 2015.
[47] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, Jul. 2017.
[48] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sanchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.
[49] S. A. Taghanaki, K. Abhishek, J. P. Cohen, J. Cohen-Adad, and G. Hamarneh, "Deep semantic segmentation of natural and medical images: A review," *Artif. Intell. Rev.*, vol. 54, pp. 1–42, Jan. 2020.
[50] H. Seo, M. B. Khuzani, V. Vasudevan, C. Huang, H. Ren, R. Xiao, X. Jia, and L. Xing, "Machine learning techniques for biomedical image segmentation: An overview of technical aspects and introduction to state-of-art applications," *Med. Phys.*, vol. 47, no. 5, pp. e148–e167, May 2020.
[51] 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.
[52] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep learning techniques for medical image segmentation: Achievements and challenges," *J. Digit. Imag.*, vol. 32, no. 4, pp. 582–596, 2019.
[53] P. Meyer, V. Noblet, C. Mazzara, and A. Lallement, "Survey on deep learning for radiotherapy," *Comput. Biol. Med.*, vol. 98, pp. 126–146, Jul. 2018.
[54] Z. Akkus, A. Galimzianova, A. Hoogi, D. L. Rubin, and B. J. Erickson, "Deep learning for brain MRI segmentation: State of the art and future directions," *J. Digit. Imag.*, vol. 30, no. 4, pp. 449–459, 2017.
[55] Z.-H. Zhou, "A brief introduction to weakly supervised learning," *Nat. Sci. Rev.*, vol. 5, no. 1, pp. 44–53, 2018.
[56] T. Eelbode, J. Bertels, M. Berman, D. Vandermeulen, F. Maes, R. Bisschops, and M. B. Blaschko, "Optimization for medical image segmentation: Theory and practice when evaluating with dice score or Jaccard index," *IEEE Trans. Med. Imag.*, vol. 39, no. 11, pp. 3679–3690, Nov. 2020.
[57] I. Monedero, "A novel ECG diagnostic system for the detection of 13 different diseases," *Eng. Appl. Artif. Intell.*, vol. 107, Jan. 2022, Art. no. 104536, doi: 10.1016/j.engappai.2021.104536.
[58] A. Çalışkan, "A new ensemble approach for congestive heart failure and arrhythmia classification using shifted one-dimensional local binary patterns with long short-term memory," *Comput. J.*, vol. 65, no. 9, pp. 2535–2546, Sep. 2022.
[59] C. Kaul, N. Pears, H. Dai, R. Murray-Smith, and S. Manandhar, "FocusNet++: Attentive aggregated transformations for efficient and accurate medical image segmentation," in *Proc. IEEE 18th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2021, pp. 1042–1046.
[60] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015* (Lecture Notes in Computer Science), vol. 9351, N. Navab, J. Hornegger, W. Wells, and A. Frangi, Eds. Cham, Switzerland: Springer, 2015, doi: 10.1007/978-3-319-24574-4_28.
[61] O. Oktay et al., "Attention U-Net: Learning where to look for the pancreas," in *Proc. 1st Conf. Medical Imag. Deep Learn.*, Amsterdam, The Netherlands, 2018, pp. 1–10.
[62] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "Ce-Net: Context encoder network for 2D medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2281–2292, Oct. 2019.
[63] 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*, vol. 11045, Sep. 2018, pp. 3–11, doi: 10.1007/978-3-030-00889-5_1.
[64] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. Wu, "UNet 3+: A full-scale connected UNet for medical image segmentation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Barcelona, Spain, May 2020, pp. 1055–1059.
[65] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, Dec. 2021.
[66] A. Süleyman, K. Fatma, and K. Yilmaz, "A new approach for congestive heart failure and arrhythmia classification using down sampling local binary patterns with LSTM," *Turkish J. Elect. Eng. Comp Sci.*, vol. 30, no. 6, pp. 2145–2164, 2022, doi: 10.55730/1300-0632.3930.
[67] A. Ghorbani, D. Ouyang, A. Abid, B. He, J. H. Chen, R. A. Harrington, D. H. Liang, E. A. Ashley, and J. Y. Zou, "Deep learning interpretation of echocardiograms," *NPJ Digit. Med.*, vol. 3, no. 1, pp. 1–10, Dec. 2020.
[68] M. Alsharqi, W. J. Woodward, J. A. Mumith, D. C. Markham, R. Upton, and P. Leeson, "Artificial intelligence and echocardiography," *Echo Res. Pract.*, vol. 5, no. 4, pp. R115–R125, 2018.
[69] F. M. Asch, N. Poilvert, T. Abraham, M. Jankowski, J. Cleve, M. Adams, N. Romano, H. Hong, V. Mor-Avi, R. P. Martin, and R. M. Lang, "Automated echocardiographic quantification of left ventricular ejection fraction without volume measurements using a machine learning algorithm mimicking a human expert," *Circulat., Cardiovascular Imag.*, vol. 12, no. 9, Sep. 2019, Art. no. e009303.
[70] N. Poilvert, "Deep learning algorithm for fully-automated left ventricular ejection fraction measurement," *J. Amer. Soc. Echocardiography*, vol. 31, no. 6, pp. 2–45, 2018.

[71] G. Carneiro, "Deep learning and data labeling for medical applications," in *Proc. 1st Int. Workshop, LABELS, 2nd Int. Workshop, DLMIA Held Conjunct MICCAI*, vol. 10008. Athens, Greece: Springer, 2016, pp. 1–11.
[72] D. G. Gungor, B. Rao, C. Wolverton, and I. Guracar, "View classification and object detection in cardiac ultrasound to localize valves via deep learning," *Proc. Mach. Learn. Res.*, London, U.K., 2020, pp. 1–12.
[73] M. Bertalmio, A. L. Bertozzi, and G. Sapiro, "Navier-Stokes, fluid dynamics, and image and video inpainting," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Dec. 2001, pp. 355–362.
[74] A. Ziller, D. Usynin, R. Braren, M. Makowski, D. Rueckert, and G. Kaissis, "Medical imaging deep learning with differential privacy," *Sci. Rep.*, vol. 11, no. 1, pp. 1–8, Dec. 2021.
[75] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.
[76] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," 2016, *arXiv:1611.05431*.
[77] C. Szegedy, S. Ioffe, and V. Vanhoucke, "Inception-V4, inception-ResNet and the impact of residual connections on learning," 2016, *arXiv:1602.07261*.
[78] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," 2016, *arXiv:1603.06937*.
[79] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 58–1929, 2014.
[80] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," 2014, *arXiv:1412.6806*.
[81] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Computer Vision* (Lecture Notes in Computer Science), vol. 8689, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham, Switzerland: Springer, 2014.
[82] G. Klambauer T. Unterthiner, A. S. Mayr, and S. Hochreiter, "Self-normalizing neural networks," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 972–981.
[83] D. A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUs)," in *Proc. 5th Int. Conf. Learn. Represent.*, 2015, pp. 1–14.
[84] P. Hu, F. Wu, J. Peng, Y. Bao, F. Chen, and D. Kong, "Automatic abdominal multi-organ segmentation using deep convolutional neural network and time-implicit level sets," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 3, pp. 399–411, 2017.
[85] S. A. Taghanaki, Y. F. 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, Oct. 2019.
[86] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Malaysia: Pearson, 2016.
[87] M. M. Hossam, A. E. Hassanien, and M. Shoman, "3D brain tumor segmentation scheme using K-mean clustering and connected component labeling algorithms," in *Proc. 10th Int. Conf. Intell. Syst. Design Appl.*, Nov. 2010, pp. 320–324.
[88] O. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D-U-Net: Learning dense volumetric segmentation from sparse annotation," in *Proc. Int. Conf. Med. Image Comput.-Assist.* Springer, 2016, pp. 424–432.
[89] L. E. Olivier, M. G. Maritz, and I. K. Craig, "Deep convolutional neural network for mill feed size characterization," *IFAC-Papers OnLine*, vol. 52, no. 14, pp. 105–110, 2019, doi: 10.1016/j.ifacol.2019.09.172.
[90] F. Chollet et al., "Keras [internet]," GitHub, 2015. [Online]. Available: https://github.com/fchollet/keras
[91] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[92] S. Nurmaini, M. N. Rachmatullah, A. I. Sapitri, A. Darmawahyuni, B. Tutuko, F. Firdaus, R. U. Partan, and N. Bernolian, "Deep learning-based computer-aided fetal echocardiography: Application to heart standard view segmentation for congenital heart defects detection," *Sensors*, vol. 21, no. 23, p. 8007, Nov. 2021.
[93] S. Nurmaini, M. N. Rachmatullah, A. I. Sapitri, A. Darmawahyuni, A. Jovandy, F. Firdaus, B. Tutuko, and R. Passarella, "Accurate detection of septal defects with fetal ultrasonography images using deep learning-based multiclass instance segmentation," *IEEE Access*, vol. 8, pp. 196160–196174, 2020.
[94] E. L. Skeika, M. R. D. Luz, B. J. T. Fernandes, H. V. Siqueira, and M. L. S. C. De Andrade, "Convolutional neural network to detect and measure fetal skull circumference in ultrasound imaging," *IEEE Access*, vol. 8, pp. 191519–191529, 2020.
[95] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626.
[96] W. Hryniewska, P. Bombiński, P. Szatkowski, P. Tomaszewska, A. Przelaskowski, and P. Biecek, "Checklist for responsible deep learning modeling of medical images based on COVID-19 detection studies," *Pattern Recognit.*, vol. 118, Oct. 2021, Art. no. 108035.

SUDHAKAR SENGAN (Member, IEEE) received the M.E. degree from the Faculty of Computer Science and Engineering, Anna University, Chennai, Tamil Nadu, India, in 2007, and the Ph.D. degree in information and communication engineering from Anna University. He is currently working as a Professor and the Director of International Relations, Department of Computer Science and Engineering, PSN College of Engineering and Technology (Autonomous), Tirunelveli. He has 20 years of experience in teaching/research/industry. He has published papers in 140 international journals, 20 international conferences, and ten national conferences. His research interests include security, MANET, the IoT, cloud computing, and machine learning. He has filed 20 Indian and three international patents in various fields of interest. He is a member of various professional bodies, such as MISTE, MIAENG, MIACSIT, MICST, MIE, and MIEDRC. He guided more than 100 Projects for UG and PG students in engineering streams. He is the Recognized Research Supervisor at Anna University under the Information and Communication Engineering Faculty. He received the Award of Honorary Doctorate (Doctor of Letters-D.LITT.) from International Economics University; SAARC Countries in Education and Students Empowerment, in April 2017. He has published three textbooks for Anna University, Chennai Syllabus.

ABOLFAZL MEHBODNIYA (Senior Member, IEEE) received the Ph.D. degree from INRS-EMT, University of Quebec, Montreal, Canada, in 2010. He is currently an Associate Professor and the Head of the ECE Department, Kuwait College of Science and Technology. Before coming to KCST, he worked as a Marie-Curie Senior Research Fellow at University College Dublin, Ireland. Prior to that, he worked as an Assistant Professor at Tohoku University, Japan, and as a Research Scientist at Advanced Telecommunication Research (ATR) International, Kyoto, Japan. His research interests include the field of communications engineering, the IoT and artificial intelligence in wireless networks, and real-world applications. He was a recipient of numerous awards, including JSPS Young Faculty Startup Grant, KDDI Foundation Grant, Japan Radio Communications Society (RCS) Active Researcher Award, European Commission Marie Skodowska-Curie Fellowship, and NSERC Visiting Fellowships in Canadian Government Laboratories. He is a Senior Member of IEICE.

SURBHI BHATIA (Member, IEEE) received the Ph.D. degree in computer science and engineering from Banasthali University, in 2018. She is currently working as an Associate Professor with the Department of Information Systems, College of Computer Sciences and Information Technology, King Faisal University, Saudi Arabia. She is also working as an Adjunct Professor at Shoolini University, India. She has more than ten years of academic and teaching experience. She has published many patents, books, and research papers in reputed journals and conferences. Her research interests include knowledge management, information systems, artificial intelligence, and data analytics. She received the Project Management Professional Certification from Reputed Project Management Institute, USA.

S. S. SARANYA (Member, IEEE) is currently an Assistant Professor with the Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur Campus, Chennai, India. She is currently pursuing the Ph.D. degree in CSE with the B. S. Abdur Rahman Institute of Science and Technology, Vandalur, Chennai. She has over 12 years of experience in teaching and research. She has published nearly ten papers in various national and international peer-reviewed journals. Her research interests include data mining, machine learning, deep learning, big data analytics, WSN, and the IoT.

MESHAL ALHARBI received the M.Sc. degree in computer science from Wayne State University, USA, in 2014, and the Ph.D. degree in computer science from Durham University, U.K., in 2020. He is currently an Assistant Professor in artificial intelligence with the Department of Computer Science, Prince Sattam bin Abdulaziz University, Saudi Arabia. He has ten years of experience in teaching/research/industry. His research interests include the artificial intelligence applications and algorithms, agent-based modeling and simulation applications, disaster/emergency management and resilience, optimization applications, and machine learning.

SHAKILA BASHEER is currently an Assistant Professor with the Department of Information Systems, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia. She has more than ten years of teaching experience and has published more technical papers in international journals/proceedings of international conferences/edited chapters of reputed publications. She has worked and contributed in the field of data mining, image processing, and fuzzy logic. Her research interest includes data mining algorithms using fuzzy logic. She is currently working on data mining, vehicular networks machine learning, block chain, vehicular networks, and the IoT.

V. SUBRAMANIYASWAMY received the B.E. degree in computer science and engineering from Bharathidasan University, India, the M.Tech. degree in information technology from Sathyabama University, India, and the Ph.D. degree from Anna University, India. He is currently working as a Professor with SASTRA Deemed University, Thanjavur, India. In total, he has 18 years of experience in academia. He continued the extension work with the Department of Science and Technology support as a Young Scientist Award Holder. He has contributed more than 150 papers and chapters for many high-quality Scopus and SCI/SCIE indexed journals and books. He is on the reviewer board of several international journals and has been a program committee member for several international/national conferences and workshops. He also serves as a guest editor for various special issues of reputed international journals. He is serving as a research supervisor and a visiting expert at various universities in India. His technical competencies lie in recommender systems, social networks, the Internet of Things, information security, and big data analytics.