Deep Learning Technique for Congenital Heart Disease Detection using Stacking-based CNN-LSTM Models from Fetal Echocardiogram: A Pilot Study

TAWSIFUR RAHMAN[1†], MAHMOUD KHATIB A. A. AL-RUWEIDI[2,3†], MD. SHAHEENUR ISLAM SUMON[1], REEMA YOUSEF KAMAL[4], MUHAMMAD E. H. CHOWDHURY[1*], HUSEYIN C. YALCIN[2*]

[1]Department of Electrical Engineering, Qatar University, Doha-2713, Qatar
[2]Biomedical Research Center, Qatar University, Doha, Qatar
[3]Department of Pharmaceutical Sciences, College of Pharmacy, QU Health, Qatar University, Doha, Qatar
[4]Pediatric Cardiology Division, Hamad General Hospital, Hamad Medical Corporation, Doha, Qatar
†. Co-first Authors
*Corresponding author: Huseyin c. Yalcin(hyalcin@qu.edu.qa), Muhammad E.H. Chowdhury (mchowdhury@qu.edu.qa)

ABSTRACT Congenital heart defects (CHDs) are a leading cause of death in infants under 1 year of age. Prenatal intervention can reduce the risk of postnatal serious CHD patients, but current diagnosis is based on qualitative criteria, which can lead to variability in diagnosis between clinicians. Objectives: To detect morphological and temporal changes in cardiac ultrasound (US) videos of fetuses with hypoplastic left heart syndrome (HLHS) using deep learning models. A small cohort of 9 healthy and 13 HLHS patients were enrolled, and ultrasound videos at three gestational time points were collected. The videos were preprocessed and segmented to cardiac cycle videos, and five different deep learning CNN-LSTM models were trained (MobileNetv2, ResNet18, ResNet50, DenseNet121, and GoogleNet). The top-performing three models were used to develop a novel stacking CNN-LSTM model, which was trained using five-fold cross-validation to classify HLHS and healthy patients. The stacking CNN-LSTM model outperformed other pre-trained CNN-LSTM models with the accuracy, precision, sensitivity, F1 score, and specificity of 90.5%, 92.5%, 92.5%, 92.5%, and 85%, respectively for video-wise classification, and with the accuracy, precision, sensitivity, F1 score, and specificity of 90.5%, 92.5%, 92.5%, 92.5%, and 85%, respectively for subject-wise classification using ultrasound videos. This study demonstrates the potential of using deep learning models to classify CHD prenatal patients using ultrasound videos, which can aid in the objective assessment of the disease in a clinical setting.

INDEX TERMS Congenital heart defects (CHDs), Fetal Echocardiogram, Deep Learning, CNN-LSTM, Stacking Machine Learning, Hypoplastic left heart syndrome (HLHS)

I. INTRODUCTION

Congenital heart defects (CHDs) account for 1% of all live births worldwide[1]. Generally, hyperplasia refers to a condition of delayed or stunted development in which an organ or a part of it remains below its normal size or remains immature[2]. Hypoplastic left heart syndrome (HLHS) is a group of cardiac malformations characterized by underdevelopment of both the aorta and the left heart, resulting in significantly impaired blood flow into the systemic circusslation and inadequate support for the circulation by the left heart[3]. HLHS is a very severe form of CHD characterized by an insufficient and non-viable left ventricle (LV) caused by congenital abnormalities that compromise the LV's ability to perform its prefusion function[4]. The incidence of HLHS is estimated to be between 0.016% and 0.036% of all live births; the occurrence is estimated to be in approximately 2 out of every 10,000 pregnancie[5, 6]. HLHS accounts for 1 to 3.8 % of congenital cardiac malformations, 8–12% of heart defects of infants with critical heart disease, and critically responsible for 25% to 40% of all neonatal cardiac mortality[7, 8].

There is currently no definitive explanation for the etiology of HLHS. Higher incidence within families with disease history suggests genetic contribution. The higher incidence of HLHS in families with a disease history suggests a genetic contribution. In some children, isolated HLHS is known to have a genetic basis. These cases may be due to mutations in the GJA1 gene with autosomal recessive inheritance or the NKX2-5 gene with autosomal dominant inheritance [9,10]. However, in the majority of the cases, the disease is diagnosed without any genetic relevance. Clinically, disturbed hemodynamics have been shown as a major contribution to the fetal development of the disease [11–13]. The growth of the left ventricle is hindered when there is a disturbance of blood flow or when the foramen ovale is affected during fetal development. Patients with HLHS have a diminution of the foramen ovale [14]. HLHS is also associated with anatomical abnormalities of the atrial septum when the superior edge of the septum and/or the primum, deviates posteriorly and leftward, resulting in obstruction of the atrial shunt[15]. The abnormal development in cardiac valves or the left ventricle itself may be caused by HLHS[16,17]. We have recently revealed evolving

IEEE Access

hemodynamics in normally and HLHS diagnosed human fetuses and demonstrated severe hemodynamic abnormalities in fetal HLHS hearts [18,19]. Animal studies supported these observations, in which surgical interventions causing blood flow abnormalities resulted in ventricular hypoplasia in the developing embryo [20–22].

The main treatment for HLHS involves a series of surgical procedures aimed at establishing the right ventricle as the main pumping chamber of the heart after birth. These procedures, collectively known as surgical palliation for HLHS neonates, involve three steps: the Norwood Procedure, the Bi-directional Glenn Operation, and the Fontan Operation [23]. These procedures establish a new functional systemic circuit in patients with HLHS. [24]. During the initial days of a newborn's life, the Norwood Procedure is conducted to establish the right ventricle as the primary pump for pulmonary and systemic circulation throughout the body. This is achieved through a connection made between the left and right atria via atrial septectomy. Subsequently, the narrowed outflow track is reconstructed by creating a connection between the right ventricle and the aorta using tissue grafts from the distal main pulmonary artery. The final step in providing pulmonary blood flow is the aortopulmonary shunt, which connects the aorta with the main pulmonary artery [25]. After a six-month recovery period following the Norwood surgery, the bidirectional Glenn procedure is performed. [26]. During this procedure, the shunt placed between the pulmonary arteries and the right pulmonary artery during the Norwood procedure is disconnected, and the right pulmonary artery is then connected to the superior vena cava (SVC). [26]. This allows for blood from the upper part of the body to enter the pulmonary artery directly, bypassing the ventricles. The Fontan operation is the third and final surgical procedure and is typically performed between 18 to 36 months after the Glenn procedure. During this procedure, a channel is created through or outside the heart to connect the vena cava to the pulmonary artery and direct blood flow to the pulmonary artery [26].

Recently, alternative surgical approaches have been proposed for treating HLHS in the fetus. One such approach is fetal valvuloplasty (FV), which is aimed at improving left heart hemodynamics, promoting growth, and maintaining biventricular circulation at birth. [27]. FV may be performed to prevent the progression of severe mid-gestation US [28]. The FV procedure involves balloon dilation inflation of the aorta to reduce fetal aortic stenosis in utero [28–35]. In a pioneering study with this approach on 100 HLHS-diagnosed fetuses and 43% of live-born patients had biventricular circulation, demonstrating the feasibility of the approach [33].

The fetal diagnosis of HLHS is of utmost importance for therapy planning, as well as for the advancement of new approaches such as fetal surgeries, as mentioned earlier. There are multiple tools available for the diagnosis of HLHS, including Computed Tomography Angiography (CTA), Cardiac Catheterization, Chest X-ray Radiography (CXR), Electrocardiography (ECG), and Echocardiography. However, all these techniques, except for Echocardiography, are difficult to apply to the fetus due to several limitations, such as invasiveness, radiation hazard, or acquisition of noise. Echocardiography, on the other hand, poses no danger to the fetus as it does not involve radiation and can provide accurate images and real-time measurements. A B-mode scan can be used to evaluate heart anatomy and ventricular position, while an M-mode and Doppler scan can be used for assessing valvular and vascular functionality [36]. The diagnosis of fetal HLHS through echocardiography relies heavily on qualitative criteria, which may lead to variations in diagnosis among clinicians. Despite this limitation, echocardiography remains a valuable tool for diagnosing fetal HLHS, as it allows for visualization of shunt flow, evaluation of the atrial and ventricular septum, assessment of vessel anatomy, and provision of functional information regarding the atrioventricular and outflow valves [37].

Recent studies by Masaaki et al. [38] introduced SONO, an architecture that employs a CNN to identify cardiac substructures and anomalies within fetal ultrasound videos. The technique involves a timeline visualization for detection likelihood and the computation of anomaly scores. The assessment focuses on cardiac structural anomalies (specifically Heart and Vessels), using area under the curve-receiver operating characteristic (AUC-ROC) analysis, and demonstrates competitive performance compared to established methods. Gangadhar et al. [39] investigated the feasibility of utilizing deep learning algorithms, specifically Artificial Neural Networks, to predict coronary artery disease at an early stage. The research aimed to enhance cardiac diagnosis and preventive measures through effective analysis of data patterns. Gonsalves et al.[40] explored the utilization of medical data for CHD prediction through Naive Bayes, Support Vector Machine, and Decision Tree ML methods, highlighting the potential of Naive Bayes probabilistic models in enhancing CHD detection.

Several computational approaches have been introduced in the medical field to advance diagnosis and therapy in severe clinical conditions. Computer-Assisted Diagnosis (CAD), for instance, has revolutionized medical image analysis, from oncology to cardiology. CAD medical image analysis has been attempted since 1965, when J.M. Prewitt and others

published papers on the use of computerized image analysis of cell images [41,42]. The introduction of Machine Learning (ML) has revolutionized the image analysis field in medicine. Deep neural networks such as Convolutional Neural Networks (CNNs) [43] and Long Short-Term Memory (LSTM) [44] and their hybrid model CNN-LSTM have shown outstanding performance in different Computer Vision problems [45]. CNN has the advantage of automatically extracting useful spatial features from the image [46,47] while LSTM is popular for extracting important temporal features [48] from a sequence of images or frames or a video. Therefore, the combination of CNN and LSTM can offer spatial and temporal feature extraction on temporally varying image or video data. Architecture combined with CNN and LSTM has been successfully used in Natural Language Processing (NLP) applications [49], Speech Recognition [50], Video Description [51], Action Recognition [52], and so on.

The importance of echocardiography in the diagnosis of CVDs is evident as it is the only imaging method that enables real-time imaging of the heart, thereby allowing for the immediate detection of various abnormalities. [53]. Combining clinician interpretation and machine learning (ML) has the potential to improve the accuracy of echocardiography by reducing inter- and intra-operator variability. In addition, ML can provide predictive information that may be too subtle for humans to detect. [54]. Some limitations of using echocardiography include the heavy reliance on the operator's experience and the qualitative interpretation of the heart's anatomical features [55]. This limitation can be addressed by integrating ML into echocardiography, which can introduce more automated and quantitative parameters. [56–59]. The datasets generated from echocardiography, particularly with the advancements in techniques such as 3D echocardiography, are often underutilized, and vast amounts of data remain uninterpreted. To bridge the gap between clinical and echocardiographic data, the introduction of ML and deep learning algorithms can be of great assistance. [60–66].

Accurate fetal diagnosis of CHDs is particularly an important area that will benefit from adopting ML approaches to advance echocardiography-based diagnosis. As in the case of HLHS, it is challenging to obtain an accurate diagnosis with conventional echocardiography approaches. Here, the structure and function of the fetal heart can be assessed through a variety of Ultrasound (US) techniques, including conventional 2-D imaging, M-mode imaging, and tissue Doppler imaging among others. However, it remains difficult to assess the fetal heart due to the involuntary movements of the fetus and its small size, in addition to some sonographers' lack of expertise in fetal

echocardiography [67]. Despite the great potential, to date, there is no study on the application of ML for the advancement of fetal diagnosis of CHDs. In this study, we are aiming to develop a deep learning technique for the automatic diagnosis of HLHS from fetal B-mode echocardiography. The main contributions of the paper are:

1. Several CNN-LSTM deep learning architectures with different state-of-the-art pre-trained encoders (such as MobileNetv2, ResNet18, ResNet50, DenseNet121, and GoogleNet) were investigated.

2. A novel CNN-LSTM architecture is used to extract spatial and temporal features from the B-mode Ultrasound during a cardiac cycle to identify the spatial and temporal abnormalities among healthy and HLHS patients.

3. Comparing the performance of CNN-LSTM architectures, a novel stacking CNN-LSTM model is developed to diagnose the HLHS early and precisely. Finally, classification results are reported for video-wise and subject-wise decisions.

II. METHODOLOGY

A. Study Population

Qatar's Women's Wellness and Research Center at Hamad Medical Corporation (HMC) is the largest fetal-maternal unit in the country, treating almost all congenital birth defects. The HMC and Qatar University Ethics Committees approved the study (HMC IRB MRC-03-17-0015). Pregnant women who were referred for an early fetal ultrasound and met the study's eligibility criteria were invited to participate. Participants in the control group provided written consent and were given an initial evaluation and two follow-up appointments during the study. The initial evaluation was conducted during the beginning of the second trimester, between 16-19 weeks of gestation, and the follow-up evaluations were conducted between 23-26 weeks and 31-34 weeks of gestation. During the same study period, women with fetuses diagnosed with congenital heart disease of the single ventricle (either hypoplastic left or right ventricle) were included in the study after obtaining verbal consent, as the initial and follow-up evaluations were part of their routine follow-up. Almost all cases of congenital heart disease in the study were referred and diagnosed during the last weeks of the second trimester. The study included 13 subjects with HLHS and 9 healthy control subjects.

B. Patient Selection Criteria

Women in the control group were deemed eligible if they had a scheduled routine fetal ultrasound

examination between weeks 16 and 18 of pregnancy.
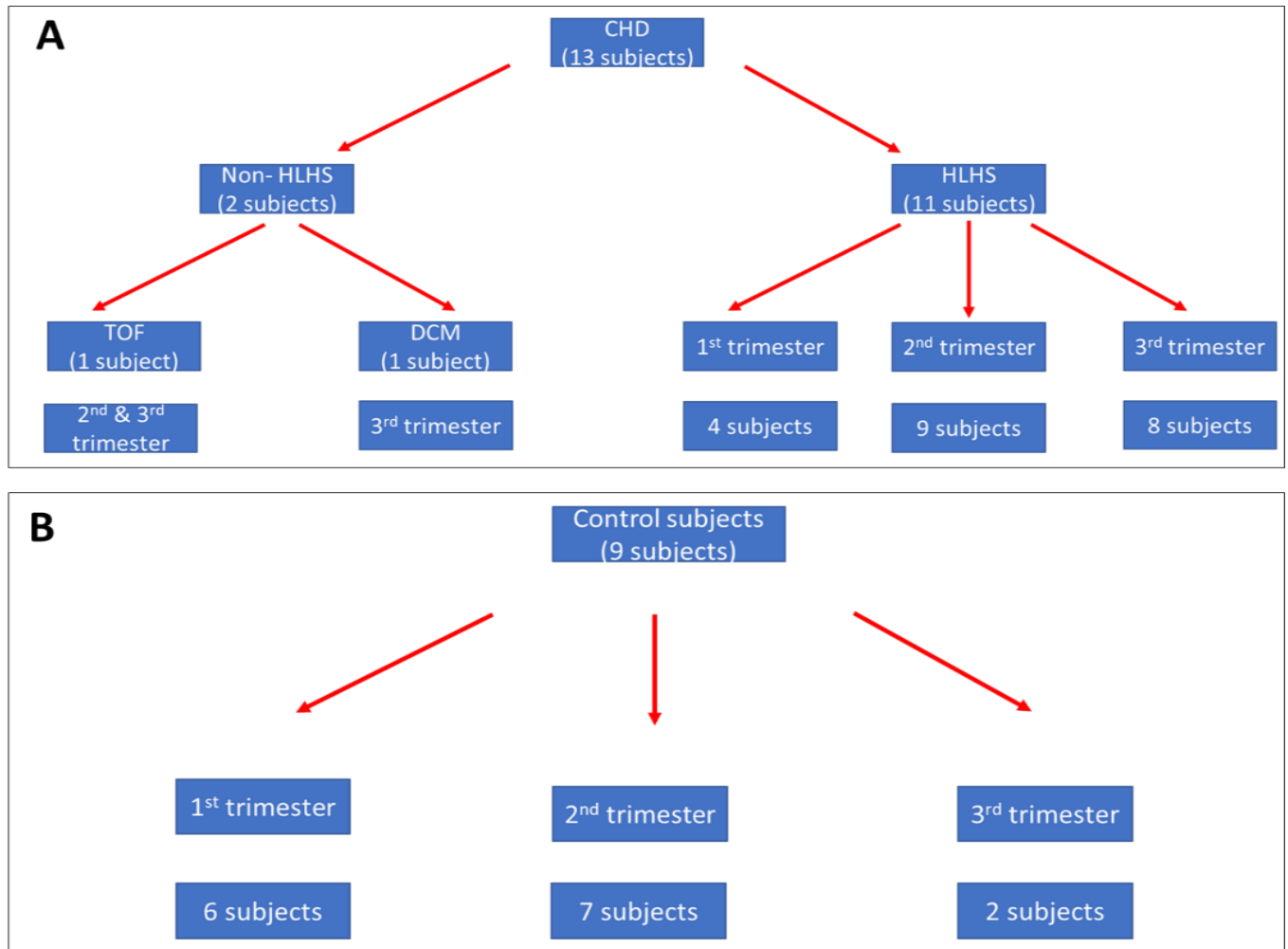
four-chamber view, diameters of the mitral and

FIGURE1 . Patient Selection Diagram.

Women referred for determining gestational age or growth discrepancy, experiencing preceding miscarriage, unable to detect a fetal heartbeat, or other miscellaneous reasons were also eligible to be part of the control group, provided the fetus was determined to be normal. Patient selection diagram is presented in Figure 1. In Figure 1(A) number of CHD subjects illustrated, trimester refers to one of the three distinct periods into which a pregnancy is divided. In Figure 1(A) number of healthy subjects illustrated.

C. Acquisition of Echocardiography Videos All examinations were conducted by a specific and experienced fetal cardiologist with ample background in fetal echocardiogram examination, using the Voluson E10 (General Electric) Ultrasound System. The Ultrasound examinations were performed with the GE RAB6-D 4D convex probe and following the guidelines issued by the American Society of Echocardiography and standards for the performance of fetal echocardiography. A detailed evaluation of all essential components of the fetal echocardiogram, including the

tricuspid valve annuli, and the lengths of the left and right ventricle, were obtained. Flow patterns across the atrioventricular and semilunar valves were evaluated using color Doppler. Doppler indices were obtained by

placing the sample volume distal to the respective valves and angled within 15-20° of the direction of blood flow. Doppler waveforms were obtained multiple times during fetal apnea, and color Doppler was used to direct the placement of the sample volume. Doppler images were used for the diagnosis of patients. However, only B-mode Ultrasound images extracted from mp4 videos were used for the ML model development in this study.
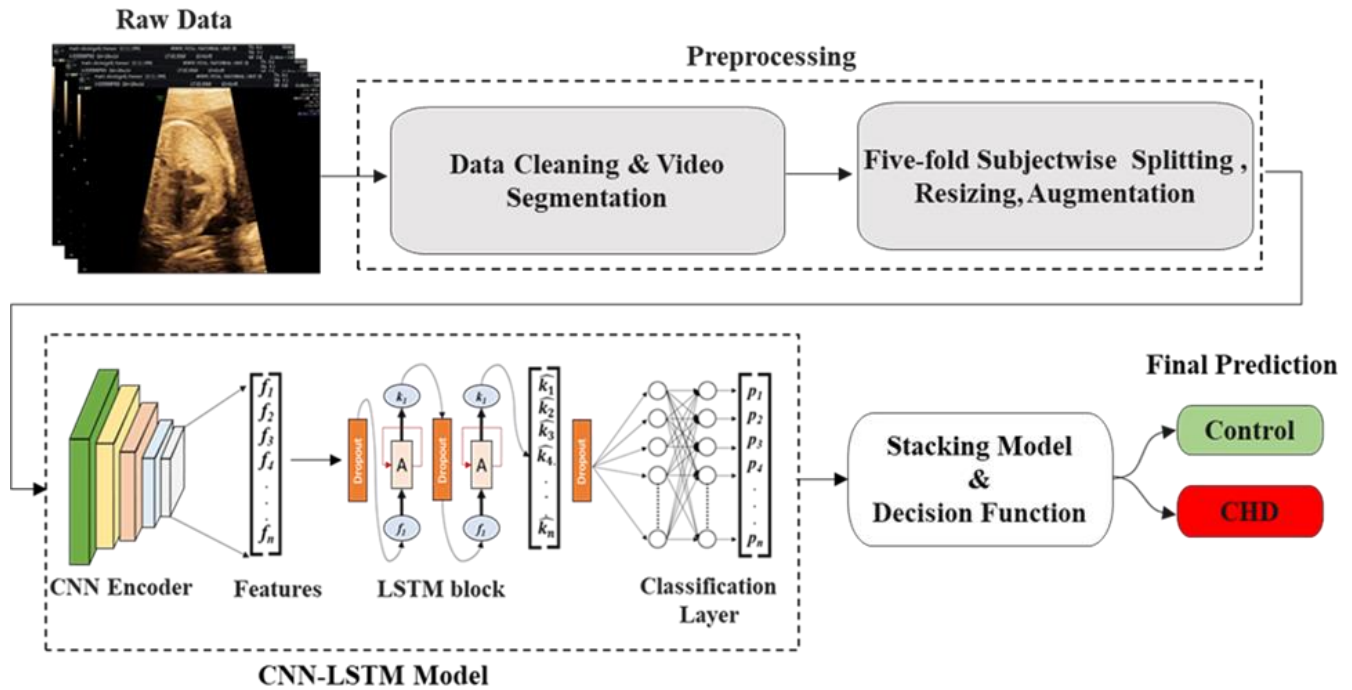
4

FIGURE2 . Overview of the methodology.

C. Machine Learning Model Development

In this study, CNN-LSTM deep learning architectures with five different encoders (such as MobileNetv2, ResNet18, ResNet50, DenseNet121, and GoogleNet) were investigated to extract useful temporal and spatial features. These features were used to train a fully connected network (FCN)/multilayer perceptron (MLP) classifier. Out of the five encoders, the three best encoder-based CNN-LSTM models' predictions were used to train a meta-classifier, a novel stacking model for the early and precise detection of CHD patients. The classification results are reported both video-wise and subject-wise. Figure 2 illustrates the schematic overview of the methodology.

D. Dataset Description

There was a total of 13 HLHS and 9 control subjects included in the study, and multiple B-mode Ultrasound videos were available for each subject in the dataset. Echocardiography was conducted at three time points: 1) 16-19 weeks gestation, 2) first follow-up at 23-26 weeks gestation, and 3) second follow-up at 31-34 weeks gestation. Each video was segmented into short videos based on the cardiac cycle. Table 1 presents the number of available videos with corresponding time points and the number of segmented videos for each patient.

E. Dataset Preprocessing

Four steps were used to process echocardiogram videos of healthy and CHD patients: 1) cleaning and cropping the videos, 2) improving the video quality, 3) segmenting the videos, and 4) training test subject-wise splitting for five-fold cross-validation for CNN-LSTM model development, validation, and testing.

1) Video Cleaning & Cropping

All static information, such as text or color, was removed from the US videos. To remove text and color from videos, firstly, we have detected fixed text and color marking regions in each frame using canny edge detection approach. Then, we replaced the identified regions with corresponding background content from the same frame or neighboring frames to achieve the removal effect. The videos were cropped by 10% from each side to eradicate unnecessary segments. Cropping eliminates unimportant elements, such as black areas, enhancing visual focus while preserving aspect ratios for uniformity. The cleaned videos were then cropped by 10% from each side to reduce unnecessary parts (black areas) of the ultrasound videos. Then, all frames of each RGB video were resized to 224×224, which is applied as input to the model. Figure 2 shows a sample of raw ultrasound frames, static information cleaned frames, and cropped frames.
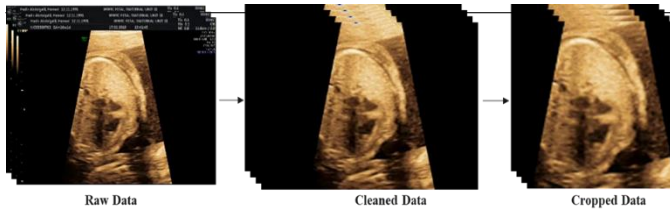
IEEE *Access*

FIGURE3 . Sample raw, cleaned, and cropped frames of the echocardiogram videos.

2) Video Enhancement

For each frame of the videos, gamma correction was applied to enhance the video quality. Typically, linear operations are performed on individual pixels in image normalization, such as scalar multiplication, addition, and subtraction [1]. In gamma correction, pixels in the source image are subjected to a non-linear operation. Gamma correction alternates the pixel value to improve the image using the projection relationship between the value of the pixel and the value of the gamma according to the internal map. If P represents the pixel value inside the [0,255] range, Ω represents the angle value, Γ is the symbol of the gamma value set, x is the grayscale value of the pixel $(x \in P)$ in equation (1-4). Let x_m be range midpoint [0, 255]. The linear map φ from group P to group Ω is defined as:

$$\varphi: P \to \Omega, \Omega = \{\omega|\omega = \varphi(x)\}, \varphi(x) = \frac{\pi x}{2x_m} \quad (1)$$

The mapping h from Ω to Γ is defined as:

$$h: \Omega \to \Gamma, \Gamma = \{\gamma|\gamma = h(x)\} \quad (2)$$
$$\begin{cases} h(x) = 1 + f_1(x) & (3) \\ f_1(x) = acos(\varphi(x)) & (4) \end{cases}$$

where $a \in [0, 1]$ denotes a weighted factor.

Based on this map h, group P can be related to Γ group pixel values. The arbitrary pixel value is calculated in relation to a given Gamma number. Let $\gamma(x) = h(x)$ and the Gamma correction function is shown in Equation 5.

$$g(x) = 255\left(\frac{x}{255}\right)^{1/\gamma(x)} \quad (5)$$

where $g(x)$ represents the output pixel correction value in the grayscale.

Table 1: Summary of the echocardiogram videos for (A) CHD subjects and (B) Healthy subjects.

CHD Subjects (A)				Healthy Subjects (B)			
Subject Id	**# of week**	**No. of full videos**	**No. of segmented videos**	**Subject Id**	**# of week**	**No. of full videos**	**No. of segmented videos**
1 (DCM)	35W	1	6	1	16W	3	51
2 (HLHS)	23W	3	27		24W	4	60
	27W	3	64	2	24W	7	87
	31W	2	21	3	25W	4	130
3 (HLHS)	28W	13	252	4	26W	3	85
	31W	2	23	5	16W	2	16
4 (HLHS)	25W	4	74		23W	4	57
5 (HLHS)	24W	5	79	6	17W	8	144
	28W	6	74		24W	4	76
	32W	2	54		33W	1	2
6 (TOF)	17W	5	104	7	19W	3	71
	18W	5	331		27W	4	57
7 (HLHS)	22W	5	88	8	19W	5	132
	25W	3	91		31W	2	36
8 (HLHS)	23W	3	80	9	17W	6	110
9 (HLHS)	24W	6	186	**Total**		**60**	**1114**
	28W	6	104				
10 (HLHS)	22W	11	168				
	26W	3	75				
	33W	8	344				
11 (HLHS)	25W	2	60				
12 (HLHS)	24W	5	266				
	32W	5	38				
	34W	3	40				
13 (HLHS)	19W	2	74				
	25W	4	46				
	30W	3	65				

Total	120	2834

| Frame 1 | Frame 2 | ················· | Frame 17 | Frame 18 | Frame 19 | Frame 20 |

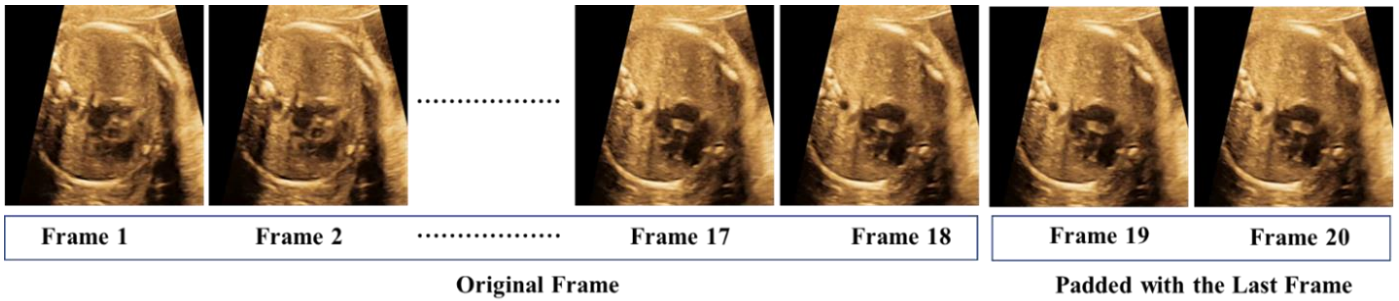**Original Frame** **Padded with the Last Frame**

FIGURE4 . Pre-processing of the Ultrasound video to up-sample to 20 frames by adding the last frame.

3) Video Segmentation

The US videos were segmented into individual heart cycle video segments using expert medical annotation. To feed these segments into the deep learning model, a fixed number of frames per video was required. The average frame-per-second (fps) of the videos are made to 33. However, some videos has much less than 20 fps while some videos have much higher than 20 frames. Therefore, each video data was pre-processed to have a maximum length of 20 frames. The choice of using 20 frames per video segment was likely determined based on a combination of factors such as the expected duration of a heart cycle, computational efficiency, and model requirements. Videos with less than 20 frames were adjusted to 20 frames by replicating the last frame, ensuring a uniform length of 20 frames across all videos. The process of video upsampling using padded frames is illustrated in Figure 4 and videos with more than 20 frames were downsampled to 20 fps.

4) Train-Test Fold Creation and Augmentation n

Five different CNN-LSTM models were investigated along with a novel stacking model.

Five-fold subject-wise cross-validation was used in this study, where 80% of the videos used for training (10% of which were used for validation) and 20% for testing. To avoid overfitting, the training data classes were made balanced since the HLHS, and healthy classes are not equal [2]. We used three popular image augmentation techniques (rotation, scaling, and translation) to make the training set balanced. Each image in the video was rotated by an angle of 5 to 10 degrees clockwise and counterclockwise for image augmentation. Each frame of the videos was scaled (magnified or reduced) by 2.5% to 10%. The images were translated horizontally and vertically by 5% to 10%. The weighted average of the five-folds is reported for each performance metric. Table 2 shows the details of the number of training, validation, and test data sets used in this study.

Table 2: Details of the dataset used for training, validation, and testing

Classes	Videos (Total Video Samples with one heart cycle)	Train video Samples (Subjects)	Train video Samples (Subjects)	Validation video Samples (Subjects)	Test video Samples (Subjects)
CHD	120 (2834)	1870 (8)	1870 (8)	404 (2)	560 (3)
Healthy	60 (1114)	750 (6)	1870 (6)	154 (1)	210(2)

F. Development of Classification Model

US videos are composed of spatial and temporal features. The spatial feature of an US video is the shape of the heart chamber, while the changes in the chamber shape during systolic and diastolic events constitute the temporal feature. In this study, CNN and LSTM layers were utilized for the extraction of spatial and temporal features, respectively. Deep CNNs have been widely employed for image classification due to their superior performance in comparison to other machine learning methods. These networks are capable of automatically extracting spatial features of an image. The approach of transfer learning has been successfully incorporated in

many applications [3-7], especially where a large dataset can be hard to find. Thus, it opens the opportunity of utilizing a smaller dataset and reduces the time required to develop a deep learning algorithm from scratch [8-9]. In this study, we used five deep learning pre-trained CNN models such as ResNet18, ResNet50 [10], DenseNet201 [11], MobileNetV2 [12], and GoogleNet [11], which were predominantly used in the literature. The feature vector after flattening the layer of CNN was fed to the LSTM layers and then it is fed to the fully connected network (FCN) or MLP for classification.

In Ultrasound videos, there is a degree of temporal connection between consecutive frames which contains

IEEE *Access*

the information about systolic and diastolic events. Neural networks such as vanilla recurrent the temporal connections in video data. LSTM [13] layers are widely used in different Ultrasound video classification [14-15]. LSTM performs better than va-

nilla RNN in this task where previous frames can preserve information in understanding the present frame [16]. LSTM layer has cell states and hidden
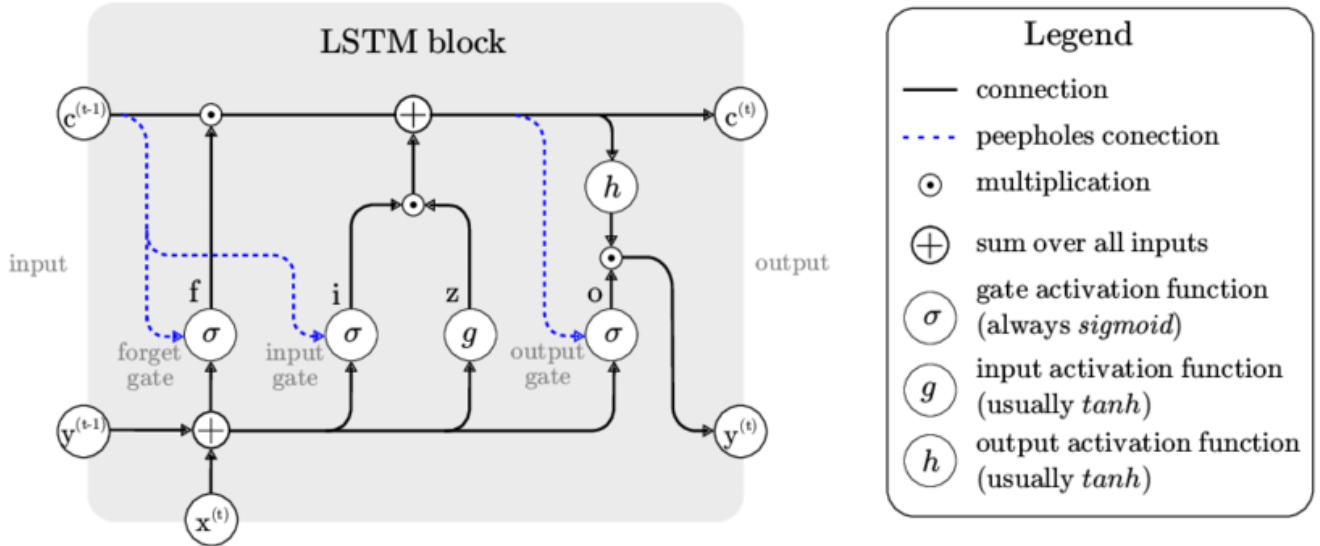

FIGURE 5. Architecture of LSTM block.

states which enables LSTM to add or remove information by regulating gates using cell states. Moreover, LSTM resolves the vanishing gradient problem of vanilla RNN by possessing the additive gradient mechanism [16]. Two layers of LSTM with 256 hidden states with a 20% dropout rate were used in this study for temporal feature extraction. Figure 5 represents the architecture of the LSTM module in the CNN-LSTM.

In this study, the stacking approach was deployed with the top-performing three CNN-LSTM models (with three different pre-trained encoders) as base learners and a

logistic regression classifier was used as meta learners to identify the CHD patients. If a single dataset A, which consists of input vectors (x_i) and their classification score (y_i). At first, a set of base-level CNN-LSTM classifiers $M_1, \ldots \ldots, M_p$ is trained and the prediction of these base learners is used to train the logistic regression-based meta-level classifier M_f as illustrated in Figure 6.
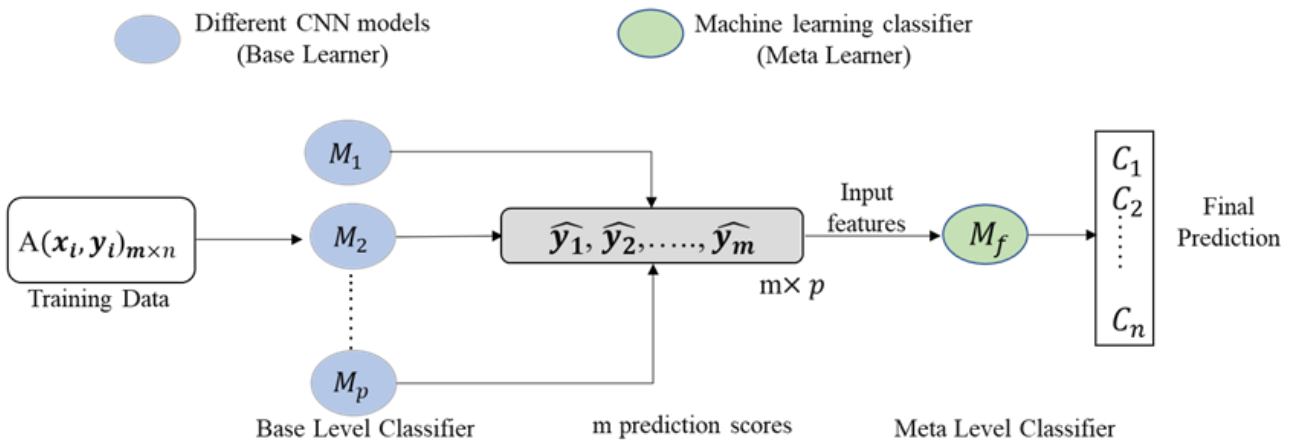

FIGURE 6. Stacking model architecture.

We used five-fold cross-validation to generate a training set for the meta-level classifier. Among these folds, base-level classifiers were trained on four-folds, leaving one-fold for testing. Each base-level classifier produces a probability value for the possible classes. Thus, using

input x, a probability distribution is created using the predictions of the base-level classifier set, M:

$$P^M(x) = (P^M(c_1|x), P^M(c_2|x), \ldots \ldots, P^M(c_n|x)) \quad (\ )$$

where $(c_1, c_2, \ldots \ldots, c_n)$ is the set of possible class values **n**, **m** denotes the number of subjects and $P^M(c_i|x)$ denotes

8

the probability that example, x belongs to a class c_j as estimated (and predicted) by the classifier, M in Equation (6). The class c_i with the highest-class probability $P^{M_j}(c_i|x)$ is predicted by a classifier, M. The metalevel classifier M_f, and attributes are thus the probabilities predicted for each possible class by each of the base-level classifiers, i.e., $P^{M_j}(c_i|x)$ for I = 1,…., n and j = 1,…., p where n, p denotes the number of classes and the number of base learners. The pseudo-code for the stacking approach is shown in Algorithm 1.

Algorithm 1: Stacking Technique

Input: training data $A = \{x_i, y_i\}_{i=1}^m$

Output: a stacking classifier M_f

1: Step 1: learn base-level classifiers
2: for t=1 to T do
3:   learn h_t based on A
4: end for
5: Step 2: construct new data set of predictions
6: for i =1 to m do
7:   $A_h = \{x_i', y_i\}$, where $x_i' = \{h_1(x_i), ……, h_T(x_i)\}$
8: end for
9: Step 3: learn a meta-classifier
10: learn M_f based on A_h
11: return M_f

G. Decision Function

A decision function was used in this study to take the decision on the final classification. The decision was taken using two different approaches: Ultrasound video-wise and subject-wise. For video-wise decisions, the average of the prediction probability scores was calculated for the segmented short videos of 20 frames individually and made one final decision for full video from all the short videos. Similarly, for subject-wise decisions, the average of the prediction probability scores of different Ultrasound videos of the subject was used to produce the final decision. Equation 6 shows the final

decision function:

$$\sigma = \frac{1}{n}\sum_{i=1}^{n} P_i(x) \qquad (7)$$

where σ is the decision function of n number of segmented or full videos. $P_i(x)$ is the probability scores for each segmented or full video.

The mean probability scores of all segmented videos were used to take the decision for the full Ultrasound video. Similarly, this study also produced the final decision of the subject by taking the mean of the probability scores of all videos of the subject

H. Experimental Setup

This study was carried out with the Pytorch package and Python 3.7. Google ColabPro was used to train all the models and the specification was 16GB Tesla T4 GPU and 120GB High RAM. Table 3 shows the training settings that were employed in this experiment.

I. Evaluation Metrics

Precision, Sensitivity, Specificity, Accuracy, F1-Score, and receiver operating characteristic (ROC) with the area under the curve (AUC), were used to evaluate the performance of different classifiers. Weighted metrics per class and overall accuracy were used as both classes had different numbers of instances. The area under the curve (AUC) was also analyzed as a metric for evaluation. Equations (8-12) show the mathematical expression of five evaluation measures (weighted sensitivity or recall, specificity, precision, overall accuracy, and F1 score):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (8) \qquad (7)$$

$$Precision = \frac{TP}{TP+FP} \qquad (9) \qquad (9)$$

$$Recall/Sensitivity = \frac{TP}{TP+FN} \qquad (10) \qquad (9)$$

$$F1\_score = 2\frac{Precision \times Sensitivity}{Precision+Sensitivity} \qquad (11) \qquad (10)$$

$$Specificity = \frac{TN}{TN+FP} \qquad (12) \qquad (11)$$

**TABLE 3.** Details of training parameters of CNN-LSTM models

Training Parameters	LSTM Hidden Size	LSTM layer	Input Dimension	Frames per video	Batch Size	Learning rate	Max Epochs	Epoch Patience	Optimizer
	256	2	224	20	4	0.0001	50	5	ADAM

Here, true positive (TP), true negative (TN), false positive (FP), and false-negative (FN) were used to denote the number of HLHS videos or subjects were identified as HLHS, the number of healthy videos or subjects were identified as healthy, the number of healthy videos or subjects incorrectly identified as HLHS and the number of HLHS videos or subjects incorrectly identified as healthy, respectively. We report the weighted performance metric, with a 95 % confidence interval, for Sensitivity, Specificity, Precision, and F1-Score, and the overall accuracy with a 95 % confidence interval for the accuracy.

III . RESULTS AND DISCUSSION

The heart is the first functional organ of the fetus. Therefore, the heart continues to function and develop at the same time. Since blood flows constantly through a developing heart, it has been suggested that hemodynamic forces (i.e. forces on cardiac tissue by flowing blood) are an important epigenetic factor governing cardiogenesis. Congenital heart disease (CHDs) form during the very complex events of heart development. These defects affect about 1% of newborn children and are the leading cause of death in infants

IEEE *Access*

under 1 year of age. CHDs can be detected prenatally via medical imaging whereas echocardiography is the most widely used technique for this purpose. Real-time imaging via echocardiography enables the assessment of heart morphology (i.e., the size of the heart chambers and valves etc.) whereas Doppler echocardiography enables the measurement of blood flow velocities through the heart (i.e., inflow through heart chambers and flow through heart valves etc.), hence evaluation of heart function. For example, prenatal echo can detect one of the most serious types of CHDs, ventricular hypoplasia (under development), as early as the 18th week of gestation, with high accuracy. Echocardiography revealed the presence of disturbed hemodynamics in hypoplastic fetal hearts, and associated abnormal forces are thought to contribute to the development of this condition. However, this evaluation is highly subjective while computer-aided-diagnosis can help here significantly. This work used echocardiography videos of a small healthy and CHD (mainly HLHS) patients cohort to develop deep learning-based detection system to automatically classify the HLHS and healthy subjects automatically and reliably. Figure 7 shows the sample of healthy and HLHS patients' 4-chamber view of the heart to show the difference in the morphology of the heart chamber in unhealthy groups during the different gestational weeks.
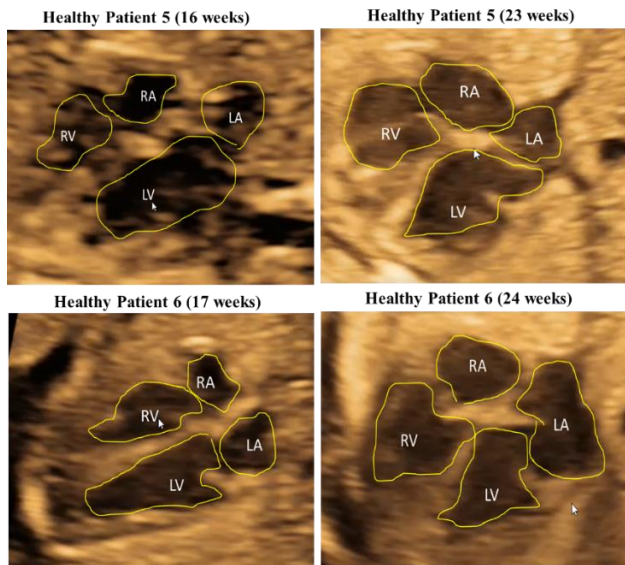
(A)
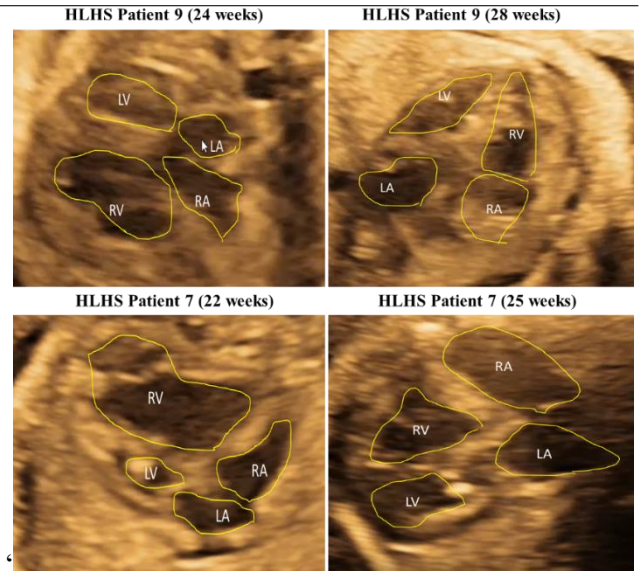
(B)

FIGURE 7. Sample snapshot for (A) healthy, and (B) HLHS patient's echocardiogram videos for two timepoints of each subject. LA is left atria, LV is left ventricle, RA is right atria, RV is right ventricle. Snapshots are for ventricle diastole in the cardiac cycle.

This study investigated and compared five different deep learning CNN-LSTM architectures using 5 pre-trained models, such as MobileNetv2, ResNet18, ResNet50, DenseNet121, and GoogleNet, for the purpose of developing a novel stacking model to predict the HLHS patients from Ultrasound videos. The results are reported using video-wise and subject-wise evaluations.

A. *Video-Wise Classification*

As discussed above, this study analyzed different deep learning LSTM models and stacking models to classify HLHS or healthy patients using echocardiogram videos. This study yielded the best performance with MobileNetv2-LSTM architecture which produced the accuracy, precision, sensitivity, F1 score, and specificity of 88.9%, 92.4%, 90.3%, 91.6%, and 86%, respectively. The stacking model was developed using the probability scores of Top-3 performing models (MobileNetv2-LSTM, ResNet18-LSTM, and GoggleNet-LSTM) which improve the result by ~2% with the accuracy, precision, sensitivity, F1 score, and specificity of 89.5%, 92.5%, 91.8%, 92.5%, and 86.2%, respectively. Table 4 shows the comparisons of different CNN-LSTM models and the stacking model for video-wise HLHS and healthy patient classification.

TABLE 4. Comparison of different CNN-LSTM performances for video-wise classification

Networks	Accuracy	Precision	Sensitivity	F1_score	Specificity
MobileNetv2+LSTM	88.9	92.4	90.3	91.6	86.0
ResNet18+LSTM	86.4	85.2	85.8	85.8	83.8
ResNet50+LSTM	83.7	84.9	85.8	85.4	83.2
DenseNet121+LSTM	82.8	82.1	84.3	83.1	82.3
GoogleNet+LSTM	84.8	84.3	85.1	84.7	84.5
Stacking Model	**90.5**	**92.5**	**91.8**	**92.5**	**86.2**

Figure 8 shows the area under the curve (AUC)/receiver-operating characteristics (ROC) curve (also known as AUROC (area under the receiver operating characteristics)) for video-wise HLHS classification using Ultrasound videos, which is one of the most important evaluation metrics for checking any classification model's performance. This is apparent from the ROC curves that the stacking CNN-LSTM model outperformed other networks for classification with 93.7% AUC whereas the AUC of the best performing CNN-LSTM (MobileNetv2-LSTM) model is 92.2%.
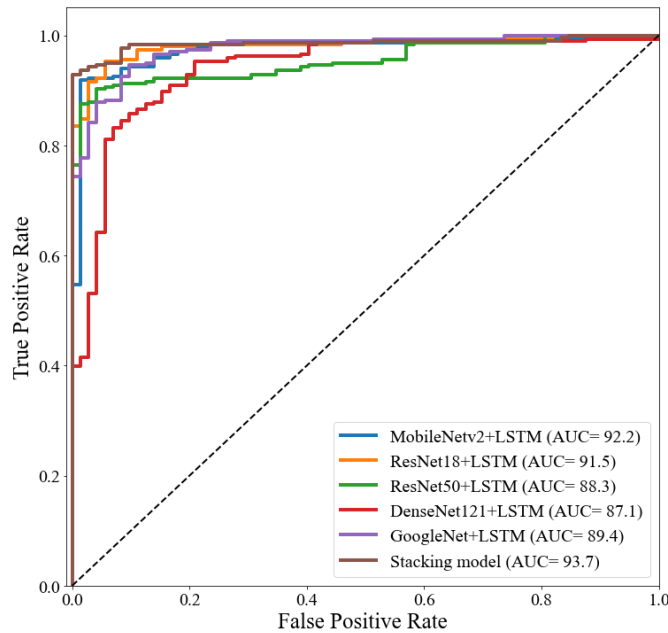

FIGURE 8. ROC curve for video-wise binary classification using different CNN-LSTM and stacking CNN-LSTM models.
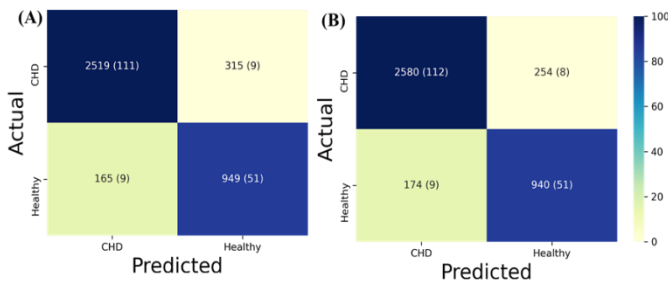

FIGURE 9. Confusion matrix for video-wise classification using (A) the best performing CNN-LSTM model, (B) and the best performing stacking CNN-LSTM model.

Figure 9 shows the confusion matrix for the best performing CNN-LSTM (MobileNetv2-LSTM) model and the stacking model for video-wise classification using echocardiogram videos. Figure 9(A) shows the confusion matrix of the best performing CNN- LSTM (MobileNetv2-LSTM) model and Figure 9(B) shows the confusion matrix of the best performing stacking CNN-LSTM model. The best performing MobileNetv2-LSTM network detects 109 videos out of 120 Ultrasound videos correctly for CHD patients while 51 videos out of 60 Ultrasound videos are correctly detected for healthy patients. On the other hand, stacking the CNN-LSTM model slightly improves the performance, where 112 videos out of 120 Ultrasound videos are detected correctly for CHD patients and 51 videos out of 60 Ultrasound videos are correctly classified as healthy.

B. Subject-Wise Classification

This study also investigated different deep learning CNN-LSTM models and stacking models for subject-wise CHD or healthy patient classification using echocardiogram videos. In subject-wise classification, this study made the decision for each subject which is the average of the predicted scores of all videos of this subject. This study produced the best performance with MobileNetv2-LSTM architecture which produced the accuracy, precision, sensitivity, F1 score, and specificity of 86.4%, 85.7%, 90.9%, 88.9%, and 80.1%, respectively. The stacking model was developed using the probability scores of Top-3 performing models (MobileNetv2-LSTM, ResNet18-LSTM, and GoogleNet-LSTM), which improved the result by ~2% with the accuracy, precision, sensitivity, F1 score, and specificity of 91%, 86.7%, 97.9%, 92.9%, and 81.4%, respectively. Table 5 shows the comparisons of different CNN-LSTM models and the stacking model for subject-wise classification.

Figure 9 shows the AUC/ROC/AUROC for subject-wise classification of HLHS and healthy subjects using Ultrasound videos. This is evident from the ROC curves that the stacking CNN-LSTM model outperformed other networks for classification with 94.5% AUC whereas the AUC of the best performing CNN-LSTM (MobileNetv2-LSTM) model is 88.4%. A significant margin of improvement in AUC (~6%) was observed in using the novel stacking technique in the final subject-wise classification.

IEEE *Access*

TABLE 5. Comparison of different CNN-LSTM performances for subject-wise classification

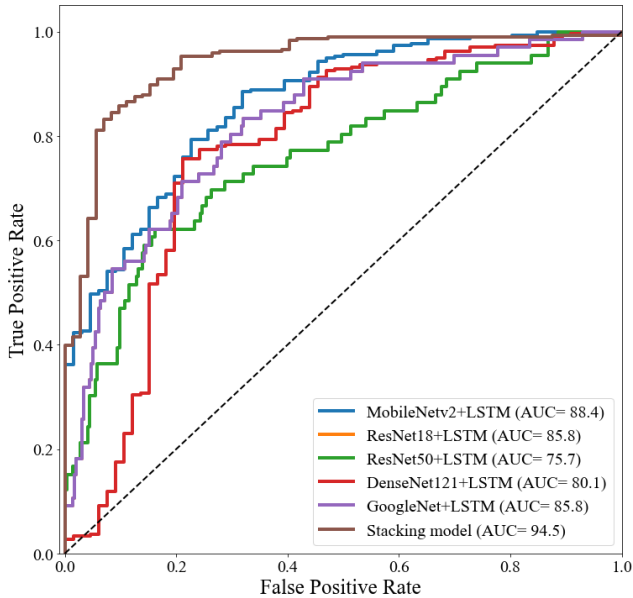| Networks	Accuracy	Precision	Sensitivity	F1_score	Specificity
MobileNetv2+LSTM	86.4	85.7	90.9	88.9	80.1
ResNet18+LSTM	84.8	80.0	90.5	85.7	75.9
ResNet50+LSTM	73.9	78.6	77.8	78.6	71.7
DenseNet121+LSTM	77.3	78.6	82.9	81.5	69.6
GoogleNet+LSTM	84.8	80.0	90.5	85.7	75.9
Stacking Model	**91.0**	**86.7**	**97.9**	**92.9**	**81.4**

FIGURE 10. ROC curve for subject-wise classification using different CNN-LSTM networks and stacking CNN-LSTM model.

Figure 11 shows the confusion matrix for the best performing CNN-LSTM (MobileNetv2-LSTM) model and the stacking model for subject-wise classification of HLHS and healthy subjects using echocardiogram videos.
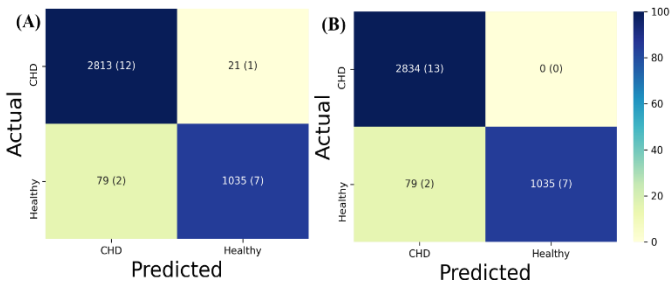

FIGURE 11. Confusion matrix for subject-wise classification using (A) the best performing CNN-LSTM model, and (B) the best performing stacking CNN-LSTM model.

Figure 11(A) shows the confusion matrix of the best performing CNN-LSTM (MobileNetv2-LSTM) model and Figure 11(B) shows the confusion matrix of the best performing stacking CNN-LSTM model. The best performing MobileNetv2-LSTM network detects 12 out of 13 subjects correctly as CHD patients while 7 out of 9 subjects are detected correctly as healthy patients. On the other hand, the stacking CNN-LSTM model improves the performance where all subjects are detected correctly as CHD patients, and 7 out of 9 subjects are correctly classified as healthy patients. The Ultrasound videos in three time-points of the two healthy subjects which were miss-classified by the model are shared as supplementary materials.

It is evident from this study that the presented novel framework that was developed using the stacking CNN-LSTM model is capable of detecting CHD patients reliably. The performance of this model can be further enhanced by increasing the sample size in the training process. For this study, there were only 9 healthy and 13 CHD patients' data available where some of the Ultrasound videos were significantly corrupted by the motion artifacts and could not be included in the analysis. Moreover, the number of videos (e.g., 60) at different time points for healthy patients was half of the CHD patients' videos (e.g, 120). Due to the limitations of healthy subjects' data, the model was not learning enough about the healthy cardiac cycle in Ultrasound videos, this might be the reason for the misclassification of the two healthy patients by the algorithm. Otherwise, the model performed outstanding in case of HLHS patient detection using Ultrasound videos. To the best of the authors' knowledge, this is the first study using Ultrasound videos to reliably classify the HLHS patients using the deep learning technique. This study can be extended with a larger patient cohort with the more longitudinal time point of Ultrasound videos to identify at which time point (i.e., gestational week) typically the deep learning model can detect the HLHS patients reliably. This will allow us to identify the more useful temporally distinctive feature(s) in the cardiac cycle of the Ultrasound videos in the different gestational week time points.

Even though, patient cohort size was low for the study, number of video samples that were used to train and test the algorithm was quite high (as shown in Table 2). For each patient, echocardiography was performed at up to three different timepoints (different gestation weeks). For a specific patient and a specific timepoint,

IEEE *Access*
Multidisciplinary : Rapid Review : Open Access Journal

several different echocardiography b-mode videos were collected from different orientations. This way, total number of full videos reached up to 180. These videos were then segmented to involve different cardiac cycles. Segmented videos were not just repetition since in most cases, fetus was moving during imaging and also operator was moving the probe for better signal. Therefore, segmented videos were treated as more sample videos in this study. The number of video samples reached to 2834 for CHD class and 1114 for healthy class which we believe, sufficient to train a deep learning model. Future work will involve testing the algorithm in larger cohorts involving different types of CHDs that are more prevalent.

III. CONCLUSION

Congenital heart disease (CHDs) affect 0.6-0.8 % of the population and are the leading cause of death in infants under 1 year of age. Current treatment of serious CHDs involves a serious of high-risk operations shortly after birth. Recently, prenatal intervention to treat these conditions has emerged as a potential therapy alternative emphasizing the importance of diagnosis of the condition in utero. Echocardiography is the gold standard for CHD diagnosis in utero. A severe CHD type is hypoplastic left heart syndrome responsible for 25% all prenatal deaths. Recently, some pioneering works showed the potential of applying ML into fetal CHD diagnosis. The objective of this study was to detect the morphological and temporal changes in the cardiac Ultrasound videos of the fetus due to HLHS using ML approaches. For this purpose, we collected echocardiography videos at different stages of gestation from fetal HLHS patients. These videos were preprocessed and used to train 5 different deep learning CNN-LSTM models and a novel stacking CNN-LSTM model. Our results suggest that the stacking CNN-LSTM model, which is developed using MobileNetv2-LSTM, ResNet18-LSTM, GoogleNet-LSTM models with different pre-trained encoders is very effective for differentiating the HLHS patients from the healthy patients using Ultrasound videos. The model could distinguish the healthy and HLHS human fetal heart differences during the gestational development stages in terms of cross-sectional heart chamber dimensions, and flow hemodynamics. This pioneering study has demonstrated that the deep learning framework is capable of distinguishing the unhealthy heart in the early gestational week using Ultrasound videos which can help in applying potential prenatal therapy rather than postnatal therapy to increase the chance of the patient survival.

SUPPLEMENTARY MATERIALS

Sample de-identified Ultrasound videos of healthy and HLHS patients are available for three acquisition time points at Supplementary Material 1 (S1). Two healthy subjects which were miss-classified by the model are shared as supplementary materials at Supplementary Material 2 (S2).

DATA AVAILABILITY

The data is available upon reasonable request.

ETHICAL COMPLIANCE STANDARDS

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The study is involving human participants and wasreviewed and approved by the Hamad Medical Corporation IRB Committee and Qatar University IRB Committee (HMC IRB MRC-03-17-0015). The patients/participants provided their written informed consent to participate in this study. The authors declare that there is no potential conflict of interest.

FUNDING

References

[1] L. Zhao, L. Chen, T. Yang, T. Wang, S. Zhang, L. Chen, Z. Ye, L. Luo, J. Qin, Birth prevalence of congenital heart disease in China, 1980-2019: a systematic review and meta-analysis of 617 studies, Eur. J. Epidemiol. 35 (2020) 631–642. https://doi.org/10.1007/s10654-020-00653-0.

[2] C.I. Tchervenkov, J.P. Jacobs, P.M. Weinberg, V.D. Aiello, M.J. Béland, S.D. Colan, M.J. Elliott, R.C.G. Franklin, J.W. Gaynor, O.N. Krogmann, H. Kurosawa, B. Maruszewski, G. Stellin, The nomenclature, definition and classification of hypoplastic left heart syndrome, Cardiol. Young. 16 (2006) 339–368. https://doi.org/10.1017/S1047951106000291.

[3] B.R. White, D.J. Goldberg, J. Rychik, Hypoplastic Left Heart Syndrome, Echocardiography in Pediatric and Congenital Heart Disease. (2021) 405–429. https://doi.org/10.1002/9781119612858.ch21.

[4] J. Rychik, Hypoplastic Left Heart Syndrome, Circulation. 130 (2014) 629–631. https://doi.org/10.1161/CIRCULATIONAHA.114.011728.

[5] C. Ferencz, J.D. Rubin, R.J. McCarter, J.I. Brenner, C.A. Neill, L.W. Perry, S.I. Hepner, J.W. Downing, Congenital heart disease: prevalence at livebirth. The Baltimore-Washington Infant Study, Am. J. Epidemiol.

IEEE *Access*

121 (1985) 31–36. https://doi.org/10.1093/oxfordjournals.aje.a113979.

[6] C.D. Morris, J. Outcalt, V.D. Menashe, Hypoplastic left heart syndrome: natural history in a geographically defined population, Pediatrics. 85 (1990) 977–983. https://doi.org/10.1542/peds.85.6.977.

[7] M.E. Abbott, Atlas of Congenital Cardiac Disease: New Edition, McGill-Queen's Press - MQUP, 2006. https://play.google.com/store/books/details?id=1FPQcXj2Q1wC.

[8] D. van der Linde, E.E.M. Konings, M.A. Slager, M. Witsenburg, W.A. Helbing, J.J.M. Takkenberg, J.W. Roos-Hesselink, Birth prevalence of congenital heart disease worldwide: a systematic review and meta-analysis, J. Am. Coll. Cardiol. 58 (2011) 2241–2247. https://doi.org/10.1016/j.jacc.2011.08.025.

[9] R.B. Hinton Jr, L.J. Martin, M.E. Tabangin, M.L. Mazwi, L.H. Cripe, D.W. Benson, Hypoplastic left heart syndrome is heritable, J. Am. Coll. Cardiol. 50 (2007) 1590–1595. https://doi.org/10.1016/j.jacc.2007.07.021.

[10] X. Liu, H. Yagi, S. Saeed, A.S. Bais, G.C. Gabriel, Z. Chen, K.A. Peterson, Y. Li, M.C. Schwartz, W.T. Reynolds, M. Saydmohammed, B. Gibbs, Y. Wu, W. Devine, B. Chatterjee, N.T. Klena, D. Kostka, K.L. de Mesy Bentley, M.K. Ganapathiraju, P. Dexheimer, L. Leatherbury, O. Khalifa, A. Bhagat, M. Zahid, W. Pu, S. Watkins, P. Grossfeld, S.A. Murray, G.A. Porter Jr, M. Tsang, L.J. Martin, D.W. Benson, B.J. Aronow, C.W. Lo, The complex genetics of hypoplastic left heart syndrome, Nat. Genet. 49 (2017) 1152–1159. https://doi.org/10.1038/ng.3870.

[11] D.W. Benson, L.J. Martin, C.W. Lo, Genetics of Hypoplastic Left Heart Syndrome, J. Pediatr. 173 (2016) 25–31. https://doi.org/10.1016/j.jpeds.2016.02.052.

[12] E.S. Selamet Tierney, R.M. Wald, D.B. McElhinney, A.C. Marshall, C.B. Benson, S.D. Colan, E.N. Marcus, G.R. Marx, J.C. Levine, L. Wilkins-Haug, J.E. Lock, W. Tworetzky, Changes in left heart hemodynamics after technically successful in-utero aortic valvuloplasty, Ultrasound Obstet. Gynecol. 30 (2007) 715–720. https://doi.org/10.1002/uog.5132.

[13] A. Mizrahi-Arnaud, W. Tworetzky, L.A. Bulich, L.E. Wilkins-Haug, A.C. Marshall, C.B. Benson, J.E. Lock, D.B. McElhinney, Pathophysiology, management, and outcomes of fetal hemodynamic instability during prenatal cardiac intervention, Pediatr. Res. 62 (2007) 325–330. https://doi.org/10.1203/PDR.0b013e318123fd3a.

[14] J.M. Kelly, C. Anderson, C.K. Breuer, The Potential Role of Regenerative Medicine on the Future Management of Hypoplastic Left Heart Syndrome, Journal of Cardiovascular Development and Disease. 9 (2022) 107. https://doi.org/10.3390/jcdd9040107.

[15] K. Suzuki, S. Doi, K. Oku, Y. Murakami, K. Mori, S. Mimori, M. Ando, Hypoplastic left heart syndrome with premature closure of foramen ovale: report of an unusual type of totally anomalous pulmonary venous

return, Heart Vessels. 5 (1990) 117–119. https://doi.org/10.1007/BF02058329.

[16] J.A. Feinstein, D.W. Benson, A.M. Dubin, M.S. Cohen, D.M. Maxey, W.T. Mahle, E. Pahl, J. Villafañe, A.B. Bhatt, L.F. Peng, B.A. Johnson, A.L. Marsden, C.J. Daniels, N.A. Rudd, C.A. Caldarone, K.A. Mussatto, D.L. Morales, D.D. Ivy, J.W. Gaynor, J.S. Tweddell, B.J. Deal, A.K. Furck, G.L. Rosenthal, R.G. Ohye, N.S. Ghanayem, J.P. Cheatham, W. Tworetzky, G.R. Martin, Hypoplastic left heart syndrome: current considerations and expectations, J. Am. Coll. Cardiol. 59 (2012) S1-42. https://doi.org/10.1016/j.jacc.2011.09.022.

[17] H.E. Salman, R.Y. Kamal, Z.M. Hijazi, H.C. Yalcin, Hemodynamic and Structural Comparison of Human Fetal Heart Development Between Normally Growing and Hypoplastic Left Heart Syndrome-Diagnosed Hearts, Front. Physiol. 13 (2022) 856879. https://doi.org/10.3389/fphys.2022.856879.

[18] H.E. Salman, R.Y. Kamal, H.C. Yalcin, Numerical Investigation of the Fetal Left Heart Hemodynamics During Gestational Stages, Front. Physiol. 12 (2021) 731428. https://doi.org/10.3389/fphys.2021.731428.

[19] H.E. Salman, M. Alser, A. Shekhar, R.A. Gould, F.M. Benslimane, J.T. Butcher, H.C. Yalcin, Effect of left atrial ligation-driven altered inflow hemodynamics on embryonic heart development: clues for prenatal progression of hypoplastic left heart syndrome, Biomech. Model. Mechanobiol. 20 (2021) 733–750. https://doi.org/10.1007/s10237-020-01413-5.

[20] M. Alser, S. Shurbaji, H.C. Yalcin, Mechanosensitive Pathways in Heart Development: Findings from Chick Embryo Studies, J Cardiovasc Dev Dis. 8 (2021). https://doi.org/10.3390/jcdd8040032.

[21] M. Alser, H.E. Salman, A. Naïja, T.D. Seers, T. Khan, H.C. Yalcin, Blood Flow Disturbance and Morphological Alterations Following the Right Atrial Ligation in the Chick Embryo, Front. Physiol. 13 (2022). https://doi.org/10.3389/fphys.2022.849603.

[22] G. Wernovsky, The paradigm shift toward surgical intervention for neonates with hypoplastic left heart syndrome, Arch. Pediatr. Adolesc. Med. 162 (2008) 849–854. https://doi.org/10.1001/archpedi.162.9.849.

[23] W.I. Norwood, P. Lang, A.R. Castenada, D.N. Campbell, Experience with operations for hypoplastic left heart syndrome, J. Thorac. Cardiovasc. Surg. 82 (1981) 511–519. https://doi.org/10.1016/S0022-5223(19)39288-8.

[24] M. Yabrodi, C.W. Mastropietro, Hypoplastic left heart syndrome: from comfort care to long-term survival, Pediatr. Res. 81 (2017) 142–149. https://doi.org/10.1038/pr.2016.194.

[25] V. Bautista-Hernandez, A. Avila-Alvarez, G.R. Marx, P.J. del Nido, Current surgical options and outcomes for newborns with hypoplastic left heart syndrome, Anales de Pediatría (English Edition). 91 (2019) 352.e1-352.e9. https://doi.org/10.1016/j.anpede.2019.10.001.

[26] M.A. Crystal, L.R. Freud, Fetal aortic valvuloplasty to prevent progression to hypoplastic left heart syndrome in utero, Birth Defects Res. 111 (2019) 389–394. https://doi.org/10.1002/bdr2.1478.

[27] S.S. Pickard, J.B. Wong, E.M. Bucholz, J.W. Newburger, W. Tworetzky, T. Lafranchi, C.B. Benson, L.E. Wilkins-Haug, D. Porras, R. Callahan, K.G. Friedman, Fetal Aortic Valvuloplasty for Evolving Hypoplastic Left Heart Syndrome: A Decision Analysis, Circ. Cardiovasc. Qual. Outcomes. 13 (2020) e006127. https://doi.org/10.1161/CIRCOUTCOMES.119.006127.

[28] W. Tworetzky, L. Wilkins-Haug, R.W. Jennings, M.E. van der Velde, A.C. Marshall, G.R. Marx, S.D. Colan, C.B. Benson, J.E. Lock, S.B. Perry, Balloon dilation of severe aortic stenosis in the fetus: potential for prevention of hypoplastic left heart syndrome: candidate selection, technique, and results of successful intervention, Circulation. 110 (2004) 2125–2131. https://doi.org/10.1161/01.CIR.0000144357.29279.54.

[29] A.C. Marshall, W. Tworetzky, L. Bergersen, D.B. McElhinney, C.B. Benson, R.W. Jennings, L.E. Wilkins-Haug, G.R. Marx, J.E. Lock, Aortic valvuloplasty in the fetus: technical characteristics of successful balloon dilation, J. Pediatr. 147 (2005) 535–539. https://doi.org/10.1016/j.jpeds.2005.04.055.

[30] L.E. Wilkins-Haug, W. Tworetzky, C.B. Benson, A.C. Marshall, R.W. Jennings, J.E. Lock, Factors affecting technical success of fetal aortic valve dilation, Ultrasound Obstet. Gynecol. 28 (2006) 47–52. https://doi.org/10.1002/uog.2732.

[31] D.B. McElhinney, A.C. Marshall, L.E. Wilkins-Haug, D.W. Brown, C.B. Benson, V. Silva, G.R. Marx, A. Mizrahi-Arnaud, J.E. Lock, W. Tworetzky, Predictors of technical success and postnatal biventricular outcome after in utero aortic valvuloplasty for aortic stenosis with evolving hypoplastic left heart syndrome, Circulation. 120 (2009) 1482–1490. https://doi.org/10.1161/CIRCULATIONAHA.109.848994.

[32] L.R. Freud, D.B. McElhinney, A.C. Marshall, G.R. Marx, K.G. Friedman, P.J. del Nido, S.M. Emani, T. Lafranchi, V. Silva, L.E. Wilkins-Haug, C.B. Benson, J.E. Lock, W. Tworetzky, Fetal aortic valvuloplasty for evolving hypoplastic left heart syndrome: postnatal outcomes of the first 100 patients, Circulation. 130 (2014) 638–645. https://doi.org/10.1161/CIRCULATIONAHA.114.009032.

[33] A.R. Prosnitz, M. Drogosz, A.C. Marshall, L.E. Wilkins-Haug, C.B. Benson, L.A. Sleeper, W. Tworetzky, K.G. Friedman, Early hemodynamic changes after fetal aortic stenosis valvuloplasty predict biventricular circulation at birth, Prenat. Diagn. 38 (2018) 286–292. https://doi.org/10.1002/pd.5232.

[34] M. Atiyah, A. Kurdi, O. Al Tuwaijry, A. Al Sahari, M. Al Rakaf, I. Babic, F. Al Habshan, Z. Alhalees, K. Al Najashi, Fetal aortic valvuloplasty: first report of two cases from Saudi Arabia, J. Cardiothorac. Surg. 15 (2020) 150. https://doi.org/10.1186/s13019-020-01195-y.

[35] C.M. Mery, P. Bastero, S.R. Hall, A.G. Cabrera, D.J. Penny, L.S. Shekerdemian, D.B. Andropoulos, C.D. Fraser Jr, C.A. Caldarone, Texas Children's Hospital Handbook of Congenital Heart Disease, Texas Children's Hospital, 2020. https://play.google.com/store/books/details?id=uaP7DwAAQBAJ.

[36] M.L. Mendelsohn, W.A. Kolman, B. Perry, J.M. Prewitt, Morphological analysis of cells and chromosomes by digital computer, Methods Inf. Med. 4 (1965) 163–167. https://www.ncbi.nlm.nih.gov/pubmed/21748889.

[37] M.L. Mendelsohn, W.A. Kolman, B. Perry, J.M. Prewitt, Computer analysis of cell images, Postgrad. Med. 38 (1965) 567–573. https://doi.org/10.1080/00325481.1965.11695692.

[38] M. Komatsu et al., "Detection of cardiac structural abnormalities in fetal ultrasound videos using deep learning," Applied Sciences, vol. 11, no. 1, p. 371, 2021.

[39] M. S. Gangadhar, K. V. S. Sai, S. H. S. Kumar, K. A. Kumar, M. Kavitha, and S. Aravinth, "Machine Learning and Deep Learning Techniques on Accurate Risk Prediction of Coronary Heart Disease," in 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), 2023: IEEE, pp. 227-232.

[40] A. H. Gonsalves, F. Thabtah, R. M. A. Mohammad, and G. Singh, "Prediction of coronary heart disease using machine learning: an experimental analysis," in Proceedings of the 2019 3rd International Conference on Deep Learning Technologies, 2019, pp. 51-56.

[41] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE. 86 (1998) 2278–2324. https://doi.org/10.1109/5.726791.

[42] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (1997) 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

[43] Z. Xu, S. Li, W. Deng, Learning temporal features using LSTM-CNN architecture for face anti-spoofing, 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR). (2015). https://doi.org/10.1109/ACPR.2015.7486482.

[44] M.E.H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M.A. Kadir, Z.B. Mahbub, K.R. Islam, M.S. Khan, A. Iqbal, N.A. Emadi, M.B.I. Reaz, M.T. Islam, Can AI Help in Screening Viral and COVID-19 Pneumonia?, IEEE Access. 8 (undefined 2020) 132665–132676. https://doi.org/10.1109/ACCESS.2020.3010287.

[45] T. Rahman, A. Khandakar, Y. Qiblawey, A. Tahir, S. Kiranyaz, S.B. Abul Kashem, M.T. Islam, S. Al Maadeed, S.M. Zughaier, M.S. Khan, M.E.H. Chowdhury, Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images, Comput. Biol. Med. 132 (2021)

IEEE *Access*

104319. https://doi.org/10.1016/j.compbiomed.2021.104319.

[46] A. Ullah, K. Muhammad, J. Del Ser, S.W. Baik, V.H.C. de Albuquerque, Activity Recognition Using Temporal Optical Flow Convolutional Features and Multilayer LSTM, IEEE Trans. Ind. Electron. 66 (2019) 9692–9702. https://doi.org/10.1109/TIE.2018.2881943.

[47] A.M. Alayba, V. Palade, M. England, R. Iqbal, A Combined CNN and LSTM Model for Arabic Sentiment Analysis, in: Machine Learning and Knowledge Extraction, Springer International Publishing, 2018: pp. 179–191. https://doi.org/10.1007/978-3-319-99740-7_12.

[48] J. Zhao, X. Mao, L. Chen, Speech emotion recognition using deep 1D & 2D CNN LSTM networks, Biomed. Signal Process. Control. 47 (2019) 312–323. https://doi.org/10.1016/j.bspc.2018.08.035.

[49] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, T. Darrell, Long-term recurrent convolutional networks for visual recognition and description, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, openaccess.thecvf.com, 2015: pp. 2625–2634. http://openaccess.thecvf.com/content_cvpr_2015/html/Donahue_Long-Term_Recurrent_Convolutional_2015_CVPR_paper.html.

[50] X. Ouyang, S. Xu, C. Zhang, P. Zhou, Y. Yang, G. Liu, X. Li, A 3D-CNN and LSTM Based Multi-Task Learning Architecture for Action Recognition, IEEE Access. 7 (undefined 2019) 40757–40770. https://doi.org/10.1109/ACCESS.2019.2906654.

[51] P.S. Douglas, M.J. Garcia, D.E. Haines, W.W. Lai, ACCF/ASE/AHA/ASNC/HFSA/HRS/SCAI/SCCM/SCCT/SCMR 2011 appropriate use criteria for echocardiography: a report of the American College of Cardiology …, Journal of the American. (2011). https://www.jacc.org/doi/abs/10.1016/j.jacc.2010.11.002.

[52] M. Alsharqi, W.J. Woodward, J.A. Mumith, D.C. Markham, R. Upton, P. Leeson, Artificial intelligence and echocardiography, Echo Res Pract. 5 (2018) R115–R125. https://doi.org/10.1530/ERP-18-0056.

[53] N. Gahungu, R. Trueick, S. Bhat, P.P. Sengupta, G. Dwivedi, Current Challenges and Recent Updates in Artificial Intelligence and Echocardiography, Curr. Cardiovasc. Imaging Rep. 13 (2020) 5. https://doi.org/10.1007/s12410-020-9529-x.

[54] P. Thavendiranathan, Z.B. Popović, S.D. Flamm, A. Dahiya, R.A. Grimm, T.H. Marwick, Improved interobserver variability and accuracy of echocardiographic visual left ventricular ejection fraction assessment through a self-directed learning program using cardiac magnetic resonance images, J. Am. Soc. Echocardiogr. 26 (2013) 1267–1273. https://www.sciencedirect.com/science/article/pii/S089

4731713005853.

[55] C. Knackstedt, S.C.A.M. Bekkers, G. Schummers, M. Schreckenberg, D. Muraru, L.P. Badano, A. Franke, C. Bavishi, A.M.S. Omar, P.P. Sengupta, Fully Automated Versus Standard Tracking of Left Ventricular Ejection Fraction and Longitudinal Strain: The FAST-EFs Multicenter Study, J. Am. Coll. Cardiol. 66 (2015) 1456–1466. https://doi.org/10.1016/j.jacc.2015.07.052.

[56] K.W. Johnson, J. Torres Soto, B.S. Glicksberg, K. Shameer, R. Miotto, M. Ali, E. Ashley, J.T. Dudley, Artificial Intelligence in Cardiology, J. Am. Coll. Cardiol. 71 (2018) 2668–2679. https://doi.org/10.1016/j.jacc.2018.03.521.

[57] F.C. Cobey, Intelligent Algorithms in Perioperative Echocardiography: A New Era, J. Am. Soc. Echocardiogr. 30 (2017) A26–A27. https://doi.org/10.1016/j.echo.2017.08.008.

[58] P.-K. Chao, C.-L. Wang, H.-L. Chan, An intelligent classifier for prognosis of cardiac resynchronization therapy based on speckle-tracking echocardiograms, Artif. Intell. Med. 54 (2012) 181–188. https://doi.org/10.1016/j.artmed.2011.09.006.

[59] F. Levy, E.D. Schouver, L. Iacuzio, F. Civaia, S. Rusek, C. Dommerc, S. Marechaux, V. Dor, C. Tribouilloy, G. Dreyfus, Performance of new automated transthoracic three-dimensional echocardiographic software for left ventricular volumes and function assessment in routine clinical practice: Comparison with 3 Tesla cardiac magnetic resonance, Arch. Cardiovasc. Dis. 110 (2017) 580–589. https://www.sciencedirect.com/science/article/pii/S1875213617300980.

[60] F.M. Asch, N. Poilvert, T. Abraham, M. Jankowski, J. Cleve, M. Adams, N. Romano, H. Hong, V. Mor-Avi, R.P. Martin, R.M. Lang, Automated echocardiographic quantification of left ventricular ejection fraction without volume measurements using a machine learning algorithm mimicking a human expert, Circ. Cardiovasc. Imaging. 12 (2019) e009303. https://doi.org/10.1161/CIRCIMAGING.119.009303.

[61] A. Narang, V. Mor-Avi, A. Prado, V. Volpato, D. Prater, G. Tamborini, L. Fusini, M. Pepi, N. Goyal, K. Addetia, A. Gonçalves, A.R. Patel, R.M. Lang, Machine learning based automated dynamic quantification of left heart chamber volumes, Eur. Heart J. Cardiovasc. Imaging. 20 (2019) 541–549. https://doi.org/10.1093/ehjci/jey137.

[62] V. Volpato, V. Mor-Avi, A. Narang, D. Prater, A. Gonçalves, G. Tamborini, L. Fusini, M. Pepi, A.R. Patel, R.M. Lang, Automated, machine learning-based, 3D echocardiographic quantification of left ventricular mass, Echocardiography. 36 (2019) 312–319. https://doi.org/10.1111/echo.14234.

[63] M.T. Nolan, P. Thavendiranathan, Automated Quantification in Echocardiography, JACC Cardiovasc. Imaging. 12 (2019) 1073–1092. https://doi.org/10.1016/j.jcmg.2018.11.038.

[64] C.D. Reddy, J. Van den Eynde, S. Kutty, Artificial

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3316719

IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

intelligence in perinatal diagnosis and management of congenital heart disease, Semin. Perinatol. (2022) 151588. https://doi.org/10.1016/j.semperi.2022.151588.

[65] P. Garcia-Canadilla, S. Sanchez-Martinez, F. Crispi, B. Bijnens, Machine Learning in Fetal Cardiology: What to Expect, Fetal Diagn. Ther. 47 (2020) 363–372. https://doi.org/10.1159/000505021.

[66] T.G. Day, B. Kainz, J. Hajnal, R. Razavi, J.M. Simpson, Artificial intelligence, fetal echocardiography, and congenital heart disease, Prenat. Diagn. 41 (2021) 733–742. https://doi.org/10.1002/pd.5892.

[67] A. Gudigar, R. U., J. Samanth, A. Vasudeva, A. A. J., K. Nayak, R.-S. Tan, E.J. Ciaccio, C.P. Ooi, P.D. Barua, F. Molinari, U.R. Acharya, Role of Four-Chamber Heart Ultrasound Images in Automatic Assessment of Fetal Heart: A Systematic Understanding, Informatics. 9 (2022) 34. https://doi.org/10.3390/informatics9020034.

[68] S. Sridevi, S. Nirmala, ANFIS based decision support system for prenatal detection of Truncus Arteriosus congenital heart defect, Appl. Soft Comput. 46 (2016) 577–587. https://doi.org/10.1016/j.asoc.2015.09.002.

[69] S. Budd, M. Sinclair, T. Day, A. Vlontzos, J. Tan, T. Liu, J. Matthew, E. Skelton, J. Simpson, R. Razavi, B. Glocker, D. Rueckert, E.C. Robinson, B. Kainz, Detecting Hypo-plastic Left Heart Syndrome in Fetal Ultrasound via Disease-Specific Atlas Maps, in: Medical Image Computing and Computer Assisted Intervention – MICCAI 2021, Springer International Publishing, 2021: pp. 207–217. https://doi.org/10.1007/978-3-030-87234-2_20.

[70] Z. Hoodbhoy, U. Jiwani, S. Sattar, R. Salam, B. Hasan, J.K. Das, Diagnostic Accuracy of Machine Learning Models to Identify Congenital Heart Disease: A Meta-Analysis, Front Artif Intell. 4 (2021) 708365. https://doi.org/10.3389/frai.2021.708365.

Authors Biograph

Tawsifur Rahman is a PhD student in Biomedical Engineering at Johns Hopkins University and working with computational pathology. He worked as a research assistant in the machine learning group at Qatar University for about 2 years. Moreover, he received his M.Sc. degree in Biomedical Engineering from the University of Dhaka (Bangladesh) as an ICT research fellow. He has published several journal papers on medical imaging. His current research interests include biomedical image and EHR data processing, machine learning, computer vision, and data science. He has received an ICT Fellowship (2019–2020) from the ICT ministry of Bangladesh for research entitled ''Driver drowsiness detection from HRV and computer vision using machine learning.
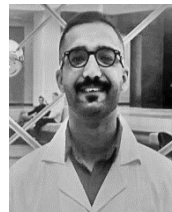
Mr. Mahmoud is a graduate student in the College of Pharmacy, Qatar University, Doha , Qatar.

Md. Shaheenur Islam Sumon, a graduate student in biomedical engineering, currently works as a Research Assssistant at the Qatar University Machine Learning Group.

Reema Yousef Kamal, a distinguished Fetal and Pediatric Cardiologist at Hamad Medical Corporation (HMC), currently serving in the esteemed Department of Pediatrics at Hamad General Hospital.

Muhammad E. H. Chowdhury (Senior Member, IEEE) received a Ph.D. degree from the University of Nottingham, U.K., in 2014. He worked as a Postdoctoral Research Fellow and a Hermes Fellow at the Sir Peter Mansfield Imaging Centre, University of Nottingham. He is currently working as an Assistant Professor with the Department of Electrical Engineering, Qatar University. He has four patents and published more than 100 peer-reviewed journal articles, conference papers, and four book chapters. His current research interests include biomedical instrumentation, signal processing, wearable sensors, medical image analysis, machine learning, embedded systems design, and simultaneous EEG/fMRI. He is currently running several NPRP, UREP, and HSREP grants from QNRF and internal grants from Qatar University, and is involved in MRC grants. He has been involved in EPSRC, ISIF, and EPSRC-ACC grants along with different national and international projects. He has worked as a Consultant for the projects entitled "Driver Distraction Management Using Sensor Data Cloud (2013–14, Information Society Innovation Fund (ISIF) Asia)." He received the ISIF Asia Community Choice Award 2013 for a project entitled "Design and Development of Precision Agriculture Information System for Bangladesh." He has recently won the "COVID-19 Dataset Award" and the "National AI Competition Awards" for his contribution to the fight against COVID-19. He is serving as an Associate Editor for IEEE Access, and a Topic Editor and a Review Editor for Frontiers in Neuroscience.

Dr. Huseyin Yalcin works as an associate research professor at Biomedical Research Center (BRC), Qatar University. He is also joint faculty for Biomedical Science and Mechanical Engineering Programs at Qatar University. His research interests include cardiovascular disorders, diabetes, and cancer focusing on basic/translational

IEEE *Access*

science aspects of the investigations through in-vitro and in-vivo systems in addition to clinical studies. Dr. Yalcin is the head of zebrafish and chick embryo facilities at BRC. Dr. Yalcin`s research experience has been fostered by several years of intensive work at state-of-the-art and multi-disciplinary institutions including Qatar Cardiovascular Research Center (as senior research scientist), and Cornell University Biomedical Engineering Department (USA, as postdoctoral researcher). He obtained his Ph.D. in Bioengineering/Mechanical Engineering in 2007 at Lehigh University, USA. Dr. Yalcin`s research interests span a wide range of topics in the basic, applied, and translational bioengineering. These involve, bioinstrumentation, biomedical signal processing, genetic manipulations, advanced microscopy imaging, ultrasonography, computed tomography, computational modeling, biomechanical characterization of tissue, and development of biomaterials for tissue engineering applications. He published more than 80 papers, some of which in top tier scientific journals like Arteriosclerosis Thrombosis and Vascular Biology, Current Biology, Biomechanics and Modeling in Mechanobiology, Frontiers in Physiology, American Journal of Physiology, Frontiers in Bioengineering and Developmental Dynamics. Dr. Yalcin has received Young Scientist award form Turkish Science Academy in 2013. Dr. Yalcin secured more than $3 million for his research funded by Qatar National Research Fund, Qatar University, European Union FP7, and Turkish Science Foundation (TUBITAK).