Ensembled Prediction of Rheumatic Heart Disease from Ungated Doppler Echocardiography Acquired in Low-Resource Settings

*****

****
****
* * * * * * * * * * * *

Abstract. Rheumatic heart disease (RHD) is a common medical condition in children in which acute rheumatic fever causes permanent damage to the heart valves, thus impairing the heart's ability to pump blood. Doppler echocardiography is a popular diagnostic tool used in the detection of RHD. However, the execution of this assessment requires the work of skilled physicians, which poses a problem of accessibility, especially in low-income countries with limited access to clinical experts. This paper presents a novel, automated, deep learning-based method to detect RHD using color Doppler echocardiography clips. We first homogenize the analysis of ungated echocardiograms by identifying two acquisition views (parasternal and apical), followed by extracting the left atrium regions during ventricular systole. Then, we apply a model ensemble of multi-view 3D convolutional neural networks and a multi-view Transformer to detect RHD. This model allows our analysis to benefit from the inclusion of spatiotemporal information and uses an attention mechanism to identify the relevant temporal frames for RHD detection, thus improving the ability to accurately detect RHD. The performance of this method was assessed using 2,136 color Doppler echocardiography clips acquired at the point of care of 591 children in low-resource settings, showing an average accuracy of 0.78, sensitivity of 0.81, and specificity of 0.74. These results are similar to RHD detection conducted by expert clinicians and superior to the state-of-the-art approach. Our novel model thus has the potential to improve RHD detection in patients with limited access to clinical experts.

Keywords: Classification, color Doppler echocardiography, deep learning, multi-view learning, rheumatic heart disease.

1 Introduction

Rheumatic heart disease (RHD), a consequence of heart valve impairment caused by acute rheumatic fever, is a common treatable condition in young children; however, its late detection and treatment can lead to heart failure or death, making it a big concern in low- and middle-income countries with limited access to specialized healthcare facilities [1], [2]. Color Doppler echocardiography is a popular test for RHD screening in children due to its safety, high speed, and cost-effectiveness [3]. In screening sonography, RHD often presents as mitral regurgitation (MR), which is blood backflow into the left atrium during ventricular systole/contraction [3]. While high-resource ultrasound devices are equipped with electrocardiogram (ECG) gating, which helps the determination of ventricular systole, hand-held ultrasound machines used in low-resource settings do not include this property.

The World Heart Federation (WHF) has established echocardiographic criteria to detect RHD based on the morphological and functional analysis of the heart valves. Using these criteria, RHD can be categorized into three groups: borderline, definite, and severe [3]. Whereas the WHF grading system is standard for RHD detection, it is complex and requires input from expert cardiologists [4], [5]. Although simplified grading systems have been proposed to detect RHD, these methods need further validation for clinical applications [6]–[8].

Imaging methods used texture analysis to assess the MR severity from echocardiograms [9], [10]. However, these methods did not detect RHD. Several other studies proposed to identify RHD from heart sound data [11], [12]. However, auscultation shows lower sensitivity than echocardiography for RHD detection [13]. Recently, convolutional neural networks (CNNs) have shown great success in several facets of automatic echocardiogram analysis [14], including view classification [15], [16], cardiac segmentation [17], and diagnosis of heart disease [18]. The state-of-the-art method for RHD detection was presented in [19]. First, a 3D CNN was used to detect RHD based on the first 16 frames of each multi-view color Doppler and B-mode echocardiogram. Then, a supervised meta-classifier was applied to aggregate the prediction results.

Recent computer vision studies have shown that fusing information from multiple views can be beneficial for improving the ability of a model to make decisions. Seeland and Mäder [20] showed that the integration of visual information through the network outperformed the fusion of classification scores by post-processing. Su et al. [21] used multiple views of a 3D object into a single deep learning model to recognize its shape. Later, Chen et al. [22] developed a multi-view vision Transformer to recognize a 3D object. Transformer, originally developed by Vaswani et al. [23], includes a self-attention mechanism in the structure of a deep learning model, which has shown great potential in many computer vision tasks [24].

We hypothesize that RHD can be accurately detected using a simplified imaging protocol based only on ungated color Doppler echocardiograms acquired at the point of care. To the best of our knowledge, the automatic detection of RHD using only color Doppler ultrasound has not been done before. In this paper, we present a complete framework for automatic RHD detection using two-view color Doppler echocardiograms (parasternal long axis [PLAXC] and apical 4-chamber [A4CC]). Our framework includes a pre-processing step for view identification, frame selection during ventricular systole, and left atrium segmentation, followed by the detection of RHD using the integration of two deep learning models (multi-view 3D CNNs and a multi-view Transformer). Our contributions include: 1) model ensemble of 3D CNNs and Transformer for the detection of RHD, 2) early fusion of visual information obtained from two views through the deep learning models, and 3) embedding spatiotemporal information with an attention mechanism to improve the accuracy of RHD detection. Early RHD detection outside of elite healthcare systems has tremendous potential to improve the quality of life even save the lives of children in low-resource settings.

2 Materials

We acquired 2,136 color Doppler echocardiograms in video format (IRB approved) from 591 children (338 females; 253 males; mean age 12±3 years; ranging from 5 to 18 years), who were examined for RHD detection. The data was acquired from at least two different views (PLAXC and A4CC) using a VIVID Q or VIVID IQ low-cost portable echocardiography machine (GE Milwaukee, WI) with a 5 MHz transducer. The data had an average image size of 597×823 pixels with a pixel resolution ranging between 0.1 and 0.4 mm. A board of expert cardiologists designated 250/591 cases as normal (no RHD) and detected RHD in 341/591 of the cases (63 definite, 260 borderline, and 18 severe RHD). For reference, the A4CC and PLAXC views, the frames during ventricular systole, and the location of the atrium were manually labeled by experts on a subset of 95 cases. Fig. 1 shows sample echocardiograms for a normal case and from patients with borderline, definite, and severe RHD.

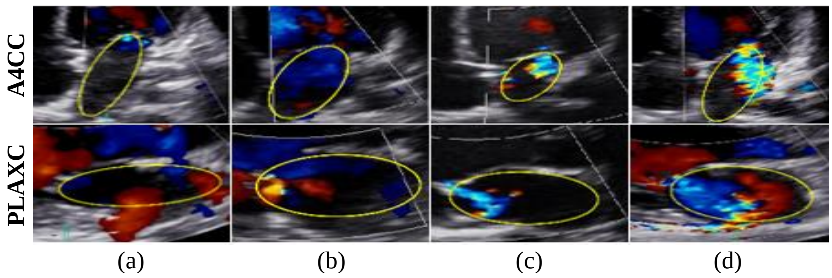

Fig. 1. Examples of the localized atrium regions (ellipsoids) on two views (A4CC and PLAXC), synchronized by the identified ventricular systole duration; (a) a normal case and patients with (b) borderline, (c) definite, and (d) severe RHD

3 Methods

Fig. 2 illustrates an overview of the proposed method, including (1) echocardiogram homogenization and (2) detection of RHD.

3.1 Echocardiogram Homogenization

Echocardiogram homogenization was performed to standardize the image information that was relevant to RHD detection, which was beneficial for improving the learning ability of models. This step included (1) view identification, (2) frame selection during ventricle contraction, and (3) left atrium segmentation.

View Identification. A4CC and PLAXC are the standard views to visualize mitral inflow/outflow for regurgitation. To retrieve these views from the multitude of collected data for each patient, we passed the first frame of each video stream (representative frame) through a deep learning classification model. The model included: (1) a ResNet-50 CNN with an input image size of 256 × 256 pixels ×3 color channels, (2) a 7 × 7 average pooling layer, (3) a fully connected layer of 512 units with rectified linear unit (ReLu) activation function, and (4) a final output layer with Softmax probability function to classify views to three categories (A4CC, PLAXC, and other). ResNet-50 CNN was pre-trained on the ImageNet datasets [25]. Parameters were selected based on the maximum accuracy criterion for the validation datasets. A size of 256 pixels was determined based on the memory required. The model was trained using the categorical cross-entropy loss function, a batch size of 32, the Adam optimization algorithm, a learning rate of 0.0001, and a total of 100 epochs.

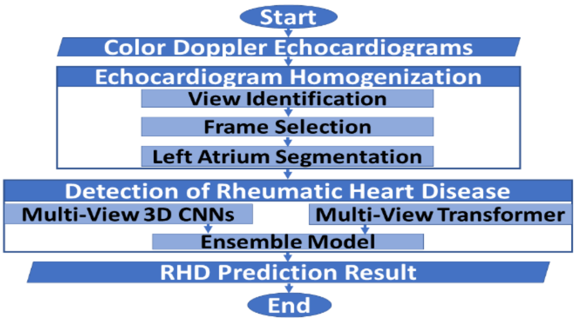

Fig. 2. Flowchart of the method proposed for RHD detection

Frame Selection during Ventricle Contraction. MR occurs during ventricle contraction/ventricular systole when the mitral valve closes. To identify frames during ventricular contraction, we employed a model with the same structure as that explained in *View Identification*, with the following notable differences: all frames from the video stream were analyzed. In addition, we used the binary cross-entropy loss function. Fig. 1 shows representative examples.

Left Atrium Segmentation. The left atrium is the region where the MR occurs. To segment the left atrium, we utilized LinkNet [26], [27] with the VGG16 encoder, pre-trained on the ImageNet datasets. In [28], LinkNet showed good accuracy for ultrasound-based kidney segmentation. The model took an input image of size 256 × 256 pixels ×3 color channels and examined it through five resolutions. All layers employed the ReLu activation function except the last one, which used the sigmoid probability function. To train the model using the reference datasets, a batch size of 32 and 500 epochs along with the Adam optimization algorithm and learning rate of 0.0001 were used to minimize the negative value of the Dice similarity coefficient as a loss function.

3.2 Detection of Rheumatic Heart Disease

We developed a model ensemble of multi-view 3D CNNs and a multi-view Transformer to fuse the spatial and temporal features with an attention mechanism for RHD detection.

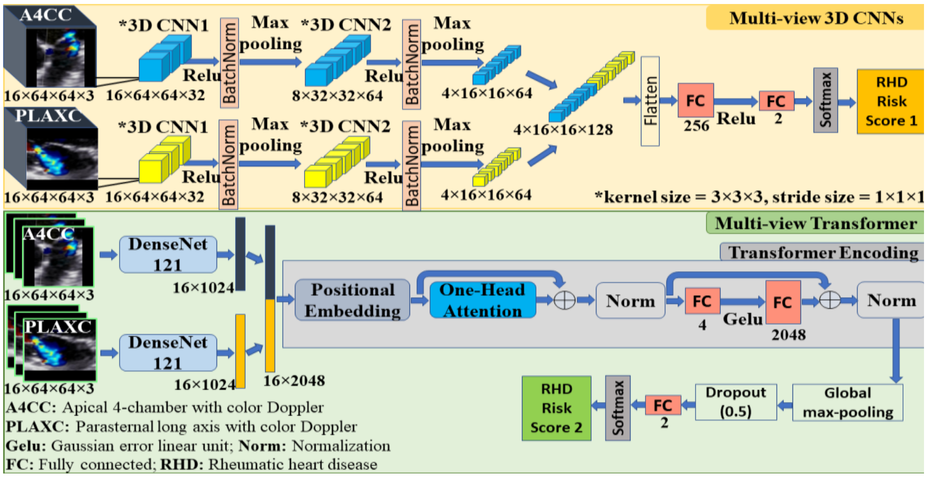

Fig. 3. The structure of the employed 3D convolutional neural networks (CNNs) and Transformer

Multi-View 3D Convolutional Neural Networks. We used 3D CNNs to extract spatiotemporal information from two views, which fused them in a single end-to-end network (Fig. 3). Input images were created from the localized left atrium regions captured during ventricular systole, resampled to 64 × 64 pixels × 3 color channels × 16 frames. These sizes were determined based on memory management. First, each 3D input image was processed using two 3 × 3 × 3 convolutional filters with the ReLu activation function. Each filter was followed by a batch normalization layer and 2 × 2 × 2 max-pooling by strides of 2 in each dimension. Then, the features extracted from the two views were concatenated and processed together using two fully connected layers, including 256 units with the ReLu activation function and 2 units with the Softmax probability function. The model employed a batch size of 64 to minimize the binary cross-entropy loss function using the Adam optimization algorithm with a learning rate of 0.0001 during 350 epochs.

Multi-View Transformer. We included a Transformer to embed relevant dependencies between the frames during ventricular contraction and draw attention to the important time points. The structure of our Transformer is shown in Fig. 3. First, we applied a DenseNet121 CNN, pre-trained on the ImageNet dataset, to capture the low-level features from the two views. DenseNet121 CNN analyzed frames during ventricular systole, resampled to 16 frames with a size of 64 × 64 pixels × 3 color channels. The information obtained from the two views was aggregated for high-level analysis by a Transformer model, similar to the one proposed in [23]. The Transformer model first embedded the positional information into the input features to keep the sequential information. Then, the model learned relationships between frames using a self-attention module followed by a feed-forward neural network, which consisted of two layers with a Gaussian error linear unit [29]. Shortcut connections between the input and output of each block were used to directly pass the gradients through the network. After each block, a normalization layer was applied to increase the generalizability and decrease the processing time. The features encoded by the Transformer model were downsampled with a global max-pooling operation and fed to the last fully connected layer with the Softmax probability function to predict RHD. Before the last fully connected layer, we applied a dropout with a keep rate of 0.5. The model was trained using a batch size of 10 and the Adam optimization algorithm (learning rate of 0.0001) to minimize the binary cross-entropy loss function through 350 epochs. Parameters were selected based on the maximum accuracy criterion for the validation datasets.

Prediction of Rheumatic Heart Disease. The 3D CNNs analyzed all frames during ventricular systole as volume data to assess RHD, while the Transformer evaluated the data frame by frame. Since the two deep learning models analyzed the data from different perspectives, we fused their predictive scores in an ensemble model by applying the maximum voting strategy to increase the accuracy. Furthermore, when multiple acquisitions were available, as is typical in clinical practice, we included all the ventricle contractions from the available A4CC and PLAXC views to obtain the final predictive score of RHD. We compared the performance of different approaches using the Wilcoxon signed-rank method with a significance level of 0.05.

4 Experimental Results

Our method was implemented using Keras (version 2.6.0) and TensorFlow (version 2.6.2) and trained on a GeForce GTX TITAN X GPU (NVIDIA, Santa Clara, CA) with 12 GB memory. We evaluated our method using cross-validation, including five folds for validation and one fold for testing. After setting aside 20% of the cases for testing, we randomly split the rest of the data into training and validation with a ratio of 80:20. Thus, RHD detection was trained on 5,108 images (378 cases), validated on 1,277 images (94 cases), and tested on 1,510 images (119 cases). Images from the same patient were not shared between the training, validation, and test sets. We randomly balanced the number of training labels for each task, including the number of (1) A4CC and PLAXC views in the view identification task, (2) frames during ventricular systoles in the frame selection task, and (3) normal and RHD cases in the RHD detection task. On a single GPU, the training time for view identification, frame selection, and atrium segmentation was 45 min, 100 min, and 250 min, respectively. The training time for RHD detection was 255 min, including 165 min and 90 min for the 3D CNN and the Transformer, respectively. While not performed in this study, parallelization is feasible.

Table 1 shows a summary of quantitative results (mean and standard deviation) for each step of homogenizing the analysis of ungated echocardiograms. Accurate performance of the identification of the image view, cardiac cycle, and atrium locations showed that our datasets were correctly homogenized, which made training easier for the RHD detection model using ungated images acquired with a manual probe. In Table 2, the quantitative RHD detection results obtained from different deep learning models and views are presented for the validation and test experiments. Integrating the information from both views increased the detection accuracy of RHD in comparison to using single views. A model ensemble of multi-view 3D CNNs and a multi-view Transformer significantly improved the performance compared to each application of 3D CNNs and Transformer (p-value of 0.03 and 0.04, respectively).

Table 1. Quantitative results obtained for each step of ungated echocardiogram homogenization

	View identification			Frame selection			Atrium localization	
Model	ResNet-50 CNN			ResNet-50 CNN			LinkNet	
View	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	DSC	HD (mm)
A4CC	0.99±0.08	1±0	0.99±0.09	0.94±0.22	0.96±0.17	0.93±0.24	0.88±0.05	0.55±0.43
PLAXC	0.99±0.08	0.99±0.07	0.99±0.08	0.93±0.23	0.94±0.23	0.93±0.23	0.9±0.04	0.45±0.22

DSC – Dice similarity coefficient; HD – Hausdorff distance.

Table 2. Quantitative RHD detection results using different deep learning models and views

	Validation							Test
Model	3D CNNs			Transformer			Ensemble	Ensemble
View	A4CC	PLAXC	Both*	A4CC	PLAXC	Both**	Both*,**	Both
Accuracy	0.68±0.37	0.71±0.35	0.73±0.35	0.67±0.41	0.68±0.36	0.7±0.36	0.75±0.34	0.78±0.41
Sensitivity	0.75±0.35	0.82±0.29	0.75±0.32	0.63±0.43	0.65±0.36	0.75±0.34	0.78±0.32	0.81±0.39
Specificity	0.59±0.37	0.56±0.37	0.71±0.31	0.72±0.36	0.73±0.35	0.63±0.36	0.72±0.33	0.74±0.44

*p-value = 0.03; **p-value = 0.04

5 Discussion

RHD is a major concern in pediatric health, especially in low- and middle-income countries with limited access to specialized clinical facilities [4]. Early detection and treatment of RHD can prevent heart failure or death. Echocardiography is an efficient exam used to diagnose RHD and trigger treatment, but it requires the input of expert cardiologists. Moreover, portable low-cost ultrasound machines used by non-experts do not use gating, which further complicates the interpretation of data. This paper proposed a fully automatic framework for the detection of RHD using only color Doppler echocardiography clips. Our approach does not require gating or ECG analysis for the detection of ventricular contraction, making it compatible with the data collected using handheld ultrasound devices. Additionally, the fully automated nature of the model will expand accessibility beyond clinical experts, allowing for RHD diagnoses to be made in areas with otherwise limited access and resources.

Assessment of color Doppler echocardiograms is challenging due, in part, to large variability in the pattern and timing of the MR jet, as well as the similarity between the spatial distribution of velocities at different locations. These challenges are also a reflection of the large variability between clinical ultrasound images acquired with manual probes. To overcome these challenges, our method homogenized the images by selecting consistent views and periods of the cardiac cycle. We focused on the left atrium regions during ventricular systole to provide the relevant information for RHD detection and reduce the variability in datasets. Since A4CC and PLAXC views assess the mitral valve from different perspectives, their information was combined through the network to provide a compact structure and improve prediction accuracy. In addition, a model ensemble of multi-view 3D CNNs and a multi-view Transformer was applied to assess the data using spatiotemporal information with an attention mechanism.

Previous reports showed that expert clinicians who reviewed echocardiograms based on the complex WHF criteria detected RHD with an agreement of 66-83% [4], [5]. Our method demonstrated a clinically acceptable accuracy of 0.78 and has the potential to extend the benefits of RHD screening without requiring the input of an expert cardiologist. In addition, our method is fully automated and reproducible. Compared to the state-of-the-art approach [19], our method detects RHD with higher accuracy (0.78 vs. 0.72) while requiring less data, i.e., only two-view color Doppler echocardiograms.

The performance of our method may be affected by low-quality frames, in particular, if the left atrium is not visible. This suggests that in the absence of experts at the point of care of the patients, minimal training should be provided to the person acquiring the data. By proposing a fully automatic framework that harmonizes images without requiring ECG gating, our method can be applied for RHD detection trials in low-resource settings. In future work, we will investigate how to assess image quality before the analysis for RHD detection.

6 Conclusion

We presented an automatic deep learning-based method to detect RHD using ungated multi-view color Doppler echocardiograms. First, we homogenized the images using deep-learning approaches to reduce data variability and to focus our classifier on the image information relevant to RHD. Next, we applied multi-view deep learning models (3D CNNs and Transformer) to analyze the spatiotemporal information of frames with an attention mechanism. Finally, we employed a model ensemble to fuse the predictions from multiple ventricular contractions and to obtain the RHD risk score. Results showed that our method could detect RHD as reliably as expert clinicians and outperformed the state-of-the-art approach for detecting RHD. Our approach is compatible with low-price, handheld ultrasound devices without ECG gating, which makes it applicable for RHD screening in low- and middle-income countries with limited access to specialists.

References

[1] Marijon, E., Mirabel, M., Celermajer, D.S., Jouven, X.: Rheumatic heart disease. Lancet 379(9819), 953–64 (2012).

[2] Liu, M., Lu, L., Sun, R.R., Zheng, Y., Zhang, P.: Rheumatic heart disease:

causes, symptoms, and treatments. Cell Biochem. Biophys. 72(3), 861–3 (2015).

[3] Reményi, B., et al.: World heart federation criteria for echocardiographic diagnosis of rheumatic heart disease--an evidence-based guideline. Nat. Rev. Cardiol. 9(5), 297–309 (2012).

[4] Beaton, A., et al.: The utility of handheld echocardiography for early rheumatic heart disease diagnosis: a field study. Eur. Hear. journal. Cardiovasc. Imaging 16(5), 475–82 (2015).

[5] Scheel, A., et al.: The inter-rater reliability and individual reviewer performance of the 2012 world heart federation guidelines for the echocardiographic diagnosis of latent rheumatic heart disease. Int. J. Cardiol. 328, 146–51 (2021).

[6] Mirabel, M., et al.: Screening for rheumatic heart disease: evaluation of a focused cardiac ultrasound approach. Circ. Cardiovasc. Imaging 8(1), 1-8 (2015).

[7] Nunes, M.C.P., et al.: Simplified echocardiography screening criteria for diagnosing and predicting progression of latent rheumatic heart disease. Circ. Cardiovasc. Imaging 12(2), 1–13 (2019).

[8] Diamantino, A., et al.: A focussed single-view hand-held echocardiography protocol for the detection of rheumatic heart disease. Cardiol. Young 28(1), 108–17 (2018).

[9] Balodi, A., Anand, R.S., Dewal, M.L., Rawat, A.: Computer-aided classification of the mitral regurgitation using multiresolution local binary pattern. Neural Comput. Appl. 32(7), 2205–15 (2019).

[10] Moghaddasi, H., Nourian, S: Automatic assessment of mitral regurgitation severity based on extensive textural features on 2D echocardiography videos. Comput. Biol. Med. 73, 47–55 (2016).

[11] Asmare, M.H., Filtjens, B., Woldehanna, F., Janssens, L., Vanrumste, B.: Rheumatic heart disease screening based on phonocardiogram. Sensors (Basel) 21(19), 1–17 (2021).

[12] Asmare, M.H., Woldehanna, F., Janssens, L., Vanrumste, B.: Rheumatic heart disease detection using deep learning from spectro-temporal representation of un-segmented heart sounds. In: Annu. Int. Engineering in Medicine and Biology Society, pp. 168–71. IEEE, Montreal, QC, Canada (2020).

[13] Godown, J., et al.: Handheld echocardiography versus auscultation for detection of rheumatic heart disease. Pediatrics 135(4), e939–44 (2015).

[14] Zamzmi, G., Hsu, L.Y., Li, W., Sachdev, V., Antani, S.: Harnessing machine intelligence in automatic echocardiogram analysis: current status, limitations, and future directions. IEEE Rev. Biomed. Eng. 14, 181–203 (2021).

[15] Madani, A., Arnaout, R., Mofrad, M., Arnaout, R.: Fast and accurate view classification of echocardiograms using deep learning. npj Digit. Med. 1(6), 1–8 (2018).

[16] Kusunose, K., Haga, A., Inoue, M., Fukuda, D., Yamada, H., and Sata, M.: Clinically feasible and accurate view classification of echocardiographic images using deep learning. Biomolecules 10(5), 1–8 (2020).

[17] Chen, C., et al.: Deep learning for cardiac image segmentation: a review. Front. Cardiovasc. Med. 7(25), 1–33 (2020).
[18] Ghorbani, A., et al.: Deep learning interpretation of echocardiograms. npj Digit. Med. 3(10), 1–10 (2020).
[19] Martins, J.F.B.S., et al.: Towards automatic diagnosis of rheumatic heart disease on echocardiographic exams through video-based deep learning. J. Am. Med. Informatics Assoc. 28(9), 1834–42 (2021).
[20] Seeland, M., Mäder, P.: Multi-view classification with convolutional neural networks. PLoS One 16(1), 1-17 (2021).
[21] Su, H., Maji, S., Kalogerakis, E., and Learned-Miller, E.: Multi-view Convolutional Neural Networks for 3D Shape Recognition. In: International Conference on Computer Vision, pp. 1–12, IEEE, Santiago, Chile (2015).
[22] Chen, S., Yu, T., Li, P.: MVT: multi-view vision transformer for 3D object recognition. In: 32nd British Machine Vision Conference, pp. 1–14, British Machine Vision Association, Online (2021).
[23] Vaswani, A., et al.: Attention is all you need. In: 31st Conference in Neural Information Processing Systems, pp. 1–11, Curran Associates Inc., Long Beach, CA, USA (2017).
[24] Khan, S., Naseer, M., Hayat, M., Zamir, S.W., Shahbaz Khan, F., and Shah, M.: Transformers in vision: a survey. ACM Comput. Surv., 1–38 (2021).
[25] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: Conference on Computer Vision and Pattern Recognition, pp. 248–55, IEEE, Miami, Florida, USA (2009).
[26] Chaurasia, A. and Culurciello, E.: LinkNet: exploiting encoder representations for efficient semantic segmentation. In: Visual Communications and Image Processing, pp. 1–5, IEEE, St. Petersburg, FL, USA (2017).
[27] Yakubovskiy, P.: Segmentation models. GitHub repository. GitHub (2019), https://github.com/qubvel/segmentation_models, last accessed 2/24/2022.
[28] Roshanitabrizi, P., et al.: Standardized analysis of kidney ultrasound images for the prediction of pediatric hydronephrosis severity. In: International Workshop on Machine Learning in Medical Imaging, Lecture Notes in Computer Science, vol 12966, pp. 366–75, Springer Cham, Strasbourg, France (2021).
[29] Hendrycks, D., Gimpel, K.: Gaussian error linear units (GELUs). arXiv: Learning, 1–9 (2016).