Communication Papers of the of the 17th Conference on Computer Science and Intelligence Systems pp. 3–9
DOI: 10.15439/2022F56
ISSN 2300-5963 ACSIS, Vol. 32

Treating Dataset Imbalance in Fetal Echocardiography Classification

Guilherme Ferreira Gusmão
Dept. of Informatics, PUC-Rio, Rio de Janeiro, Brazil
Email: gusmaof@tecgraf.puc-rio.br

Renato Cherullo de Oliveira
Tecgraf Institute, PUC-Rio, Rio de Janeiro, Brazil
Email: cherullo@tecgraf.puc-rio.br

Carlos Roberto Hall Barbosa
Postgraduate Programme in Metrology, PUC-Rio, Rio de Janeiro, Brazil
Email: hall@puc-rio.br

Alberto Barbosa Raposo
Dept. of Informatics, PUC-Rio, Rio de Janeiro, Brazil
Email: abraposo@inf.puc-rio.br

Abstract— **Deep learning has been a trending topic during the last few years, notably in medical imaging that employs neural networks for image manipulation, computer-aided detection of diseases, and many other tasks depending on the clinical practices. One possible application that would benefit from these methods is the fetal cardiac view classification, where these different views are useful to obtain valuable information about the patient's heart development. A trained network could help reduce variance in interpretation and speed up data annotation. Alas, in this context we can face two challenges: datasets may contain a lot of information not relevant to the outcome of the classifier's training, and the view classes may be unbalanced in the sense that certain classes may have much more samples than others. This paper presents a series of attempts to solve these issues and can be used as a practical guide for training viable classifiers in this context.**

Index Terms—fetal echocardiography, cardiovascular imaging, image processing, machine learning, data augmentation.

I. Introduction

DEEP learning has found a lot of traction in recent years, especially focusing on automating big data analysis. With the rise of new technologies, like more powerful GPU allowing faster parallel processing of data, many new applications can be explored [1]. One field that is taking advantage of these developments is Medical Imaging. It uses deep neural networks for image reconstruction, enhancement, segmentation, computer-aided detection of diseases, and many other tasks depending on the clinical practices, such as chest, neuro, cardiovascular, abdominal, and microscopy imaging. In this paper, we will be focusing on echocardiography tasks [2-4].

Echocardiography is the mainstay of cardiovascular imaging, as it is a fast method for image acquisition that avoids the use of radiation. The method produces grayscale videos containing sufficient information to diagnose alterations in the development of the patient's heart even in the gestational period [2]. The interpretation of the resulting images is typically made manually by a cardiologist, and it could take a lot of time and effort. To improve this method, many researchers are focusing on transforming this process into a partially automated pipeline for interpreting cardiovascular imaging using deep neural networks.

One difficult task that would benefit from deep learning methods is the fetal cardiac view classification. These views are standardized imaging planes defined in the medical literature, which cardiologists use to obtain valuable information about the development of the different valves and chambers of the heart. A trained network could help reduce variance in interpretation and speed up data annotation [5-8]. Another noteworthy use would be to find a specific heart view among the frames of a fetal echocardiography video. This is very important for the detection of congenital heart diseases that are still being missed during diagnostics [9-10]. This analysis may be made by the cardiologist through the appraisal of the four standard heart views:

- Four-chamber (4CH) view that contains aorta descendens, left atrium, left ventricle, right atrium, and right ventricle;
- Left Ventricular Outflow Tract (LVOT) view that contains aorta ascendens, left atrium, left ventricle, right atrium, and right ventricle;
- Right Ventricular Outflow Tract (RVOT) view that contains ductus arteriosus, superior vena cava, aorta ascendens, and main pulmonary artery; and
- Three Vessels View (3VV) that contains ductus arteriosus, superior vena cava, aorta ascendens. The 3VV can have a few variations, for example, in our dataset we also have the Three Vessels and Trachea (3VT) view.

Funding: This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001.

Acknowledgments: The authors thank for the financial support provided by the Brazilian funding agencies CNPq, CAPES, FINEP, and FAPERJ.

These five heart views are exemplified in Fig. 1.

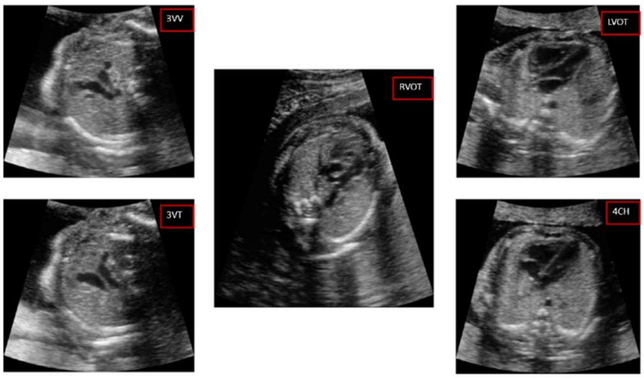

Fig. 1. The four standard fetal heart views: 3VV (+ 3VT), RVOT, LVOT and 4CH.

However, in this context we face two challenges: first, echocardiography datasets contain a lot of information that may or may not influence the outcome of the classifier's training, and there is not a default annotation format to follow. The second is that fetal echocardiography datasets typically have a large imbalance, in the sense that certain classes contain many more samples than others, as is the case of the 4CH view in relation to the other more complex views, as it is the starting point and most important view in fetal echocardiography [5-8]. Meanwhile, other views may be used only when investigating rare, specific conditions.

Unbalanced datasets are a very common issue in deep learning, not only in medical imaging, and many researchers tried to reduce its impact during training [11-14].

In [11], the authors developed a self-contained deep architecture with the synthetic minority oversampling technique (SMOTE) for artificial instance generation that doesn't need a discriminator during the synthetic image generation process, instead depending only upon a penalty function for improving the generator.

The authors of [12] proposed a generative adversarial network (GAN) with a three-player adversarial game: a generator that produces synthetic images of the minority classes to fool both the discriminator of real/fake images and a multi-class classifier. This min-max game, in time, will adjust all players for better outputs, reducing misclassifications of the underrepresented classes.

In [13], a training technique called ReMix leverages batch resampling, instance mixing and soft-labels to induct robust deep models from imbalanced and long-tailed datasets.

The RetinaNet [14] tries to solve the problem of extreme foreground-background class imbalance encountered during training of dense detectors by reshaping the standard cross entropy loss to downplay the more prevalent classes.

While most of the other works focus on the models' architectures and loss function, in this paper we tackle these imbalance challenges for training viable classifiers with a more data-centric approach, as our primary effort being on parsing, pruning, and organizing the fetal echocardiography dataset, while using readily available convolutional neural networks.

This manuscript is structured as follows: Section II presents the dataset characteristics and a brief mention of the model used in our experiments; the methods for balancing the dataset and their impact upon training are shown in Section III; conclusions of the manuscript, including recommendations for future works, are described in Section IV.

II. Dataset and Model

Deep learning applications have four fundamental cornerstones: the layers that build the network; the dataset, with the training and ground truth data; the loss function, which defines how well modeled the network is; and the optimizer, which reduces overall loss by helping better define the model weights [15].

The literature shows that we have reached a high level of maturity in the development of effective neural networks, optimizers, and loss functions, allowing for easy implementation and fine-tuning of the model hyperparameters. However, it is important to do more dataset-driven explorations [15].

Usually, the *modus operandi* for neural networks is to train them with the largest number of samples possible, but, sometimes, we have only a small collection of relevant data to work with [15]. This is especially true in medical imaging when analyzing rare occurrences or rarely performed exams. It is still possible to successfully train a network with relatively little data if all samples are representative and correctly annotated.

The dataset used for training our neural network model comes from private data supplied by another project in development at the Tecgraf Institute. It originally contained a set of Excel tables and DICOM files (Digital Imaging and Communications in Medicine) but it has been already pre-processed into a single, consolidated Excel file describing the results of each exam and referencing a set of gray-scale PNG (Portable Network Graphics) files, each containing a single frame of an exam. Although it is possible to figure out if two images belong to the same exam or even if they belong to the same patient, it is worth noting that all data is anonymized.

These exams were acquired using different machines and were annotated by different doctors. Each row contains a bit of information about either a point, a line, a region of a frame, a whole frame, a range of frames, or all the frames in an exam video. This means that there can be many rows related to a single exam frame, and some exam frames may not have any kind of annotation associated with them at all.

To evaluate the dataset's class balance, we selected all the frames annotated with view type information, and belonging to the classes 3VT, 3VV, 4CH, LVOT, and RVOT, producing the histogram presented in Fig. 2.

Right off the bat, we can see how unbalanced the dataset is: there are around forty-five times more frames in the most represented class than in the least represented class. We will tackle this imbalance in the following sections.

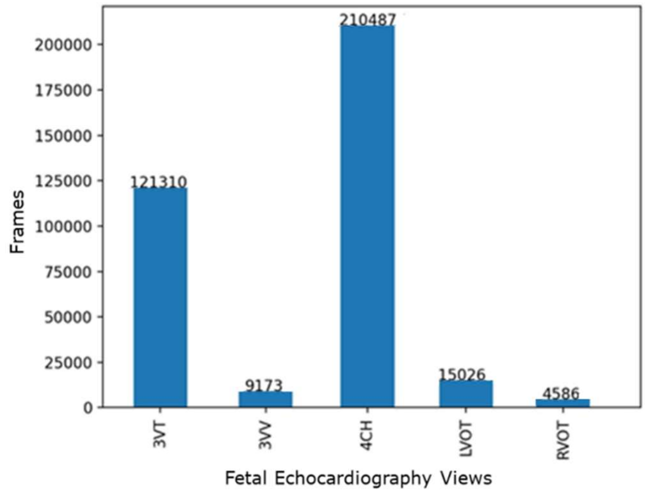

Fig. 2 Histogram with classes distribution throughout the dataset.

Inspecting the pixel value range in each image, that is, the value of the brightest pixel minus the value of the darkest pixel in each image, we produced the histogram in Fig. 3 that shows the images' pixel value range, considering that the value 1.0 corresponds to white and that the value 0.0 corresponds to black.

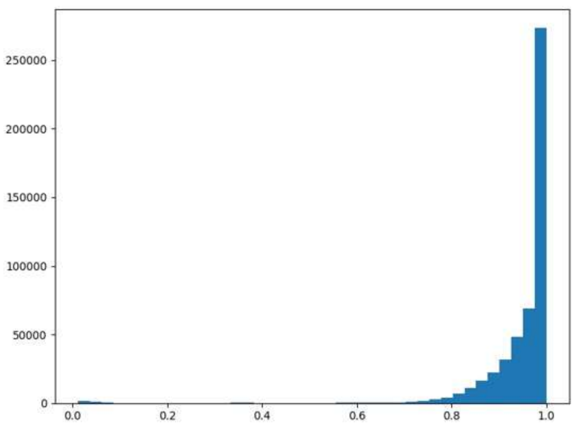

Fig. 3 Histogram of the pixel value range in each image.

Although it's not perfect, most of the images have a pixel value range above 0.8, so simply adjusting for range should not have much impact on the classifier's performance.

For our classifier model, we chose the EfficientNets [16] family of models. It is a convolutional neural network (CNN) and scaling method that uniformly scales all dimensions of depth/width/resolution using a compound coefficient, a user defined parameter that manages the resources availability, optimizing the network. We choose this architecture due to a few reasons: its baseline model has good performance and low resource requirements; it can be easily scaled up if needed; it's flexible since it can run on Tensorflow 1.15, and it was known to work to classify fetal echocardiograms since it was formerly used in a previous project at the Tecgraf Institute.

Finally, we can start rebalancing our dataset to better fit our training, as described in the next section.

III. Balancing the Data: Preprocessing and Augmentation

As seen in Fig.2, the dataset is very unbalanced. Therefore, to counteract this, many methods found throughout the literature were tested. Since the goal of this paper is to present guidelines, we decided to present our findings in a streamlined manner: instead of showing the results of each dataset preprocessing or augmentation technique individually and additively, we packed those techniques in thematic, coherent sets.

We created a baseline test using the first 8 % of the dataset for training and used the last 2 % for validation, preserving the class imbalances and without using any enhancement. We used a fraction of our data to reduce the time taken during training, so we could quickly iterate towards a working solution. It also prevents data leakage, when samples in the training set are equal or extremely similar to samples in the validation set, since the dataset was sorted by patient.

The class distributions in both sets are shown in Fig. 4.

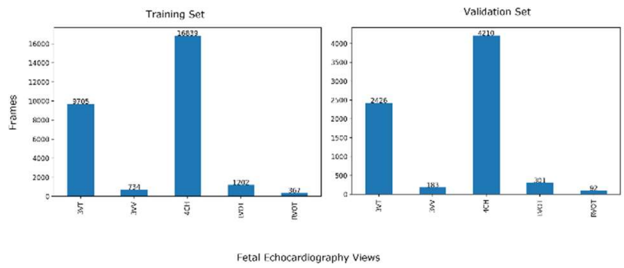

Fig. 4 Class distribution in the training and validation sets. Both keep the imbalance of classes in the dataset.

We used for training the EfficientNet B0 model with Adam as the optimizer with learning rate 0.001 (well known for its great performance in image classification problems). We reduced the learning rate on plateau monitoring the validation loss, with patience as 7, factor as 0.5 and minimum learning rate as 10^{-9}. All these values were found through empirical experiments that provided the best results for training the model. Categorical cross-entropy was chosen, as our output has five possible outcomes. EfficientNet allows you to start training with pre-trained weights from ImageNet, but we opted for training without it. Beyond that, the EfficientNet family employs a softmax activation function, so the dataset labels were codified by one-hot encoding [16].

Looking at the baseline learning history, in Fig. 5, we can see that the network quickly overfits the training set with no further improvements to the validation set loss. Whereas the confusion matrix over the validation set, in Fig. 6, shows that most of the achieved accuracy was due to the 3VT and 4CH classes (the most represented ones). The LVOT class got a F1-score of 0.11 and the network didn't predict a single image in the RVOT and 3VV classes. The weighted average F1-score was 0.54, as seen in Table I.

Things look dire for our dataset with these results, but the following subsections will bring techniques that will better improve the data quality.

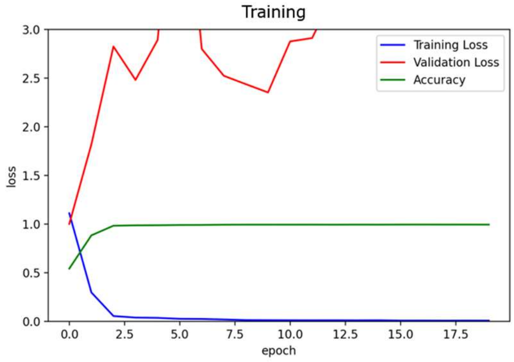

Fig. 5 Learning history with accuracy, training loss, and validation loss.

Fig. 6 Baseline Model's Confusion Matrix.

A. Preprocessing Selection

Stride

Echocardiogram exams are video data, but we opted, as suggested in other papers [5-6], to develop our classifier with still images sampled from the video. Consequently, in earlier experiments, we sampled all the images from each successive exam in a class, up to the desired frame count for that class. But, since consecutive video frames change very little from one another, the network was being fed with many similar frames, reducing the learning potential. So, instead of sampling frames contiguously in an exam, we tried sampling frames with a stride. In our case, one frame was selected every 5 frames. This brought an increase in data diversity, as a greater volume of exams from a variety of patients were sampled.

Shuffle

First and foremost, never forget: always shuffle your data after every epoch! This will help escape bad batches and minimize variance, increasing the generalization of the model for unseen data. Disregarding this truth can hinder the advances in your research [15].

So, after striding, we shuffled the entire dataset before selecting 8 % of samples for training and 2 % for validation, avoiding shared frames between the sets and once again preserving the class distributions.

This step greatly increased the diversity of the samples in each class again, by using frames from more exams and avoiding any bias that could be present at the beginning or the tail-end of the original dataset.

Stratified Batch

Another interesting technique used in unbalanced scenarios is to do what is called stratified batching [17]. It consists in, instead of feeding all images of the training set to the network every epoch, randomly choosing a fixed number of images of each class, so that at the end of each epoch, the network will have been trained with the same number of images from each class. This prevents bias toward predominant views, as it reduces the number of samples from the same patient to be used. These batches must have their size limited by the total amount of samples of the least prevalent class. In our case, we used batches of size 300.

New training was implemented with these three methods to better understand their impact on the model performance. In Fig. 7 and Fig. 8, the learning history and confusion matrix can be investigated.

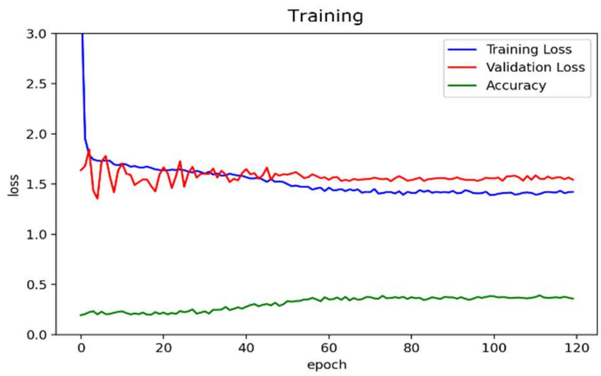

Fig. 7 Learning history after shuffling, striding, and stratified batching on the dataset.

TABLE I.
BASELINE TEST REPORT

Class	Precision	Recall	F1-score	Number of Samples
3VT	0.43	0.42	0.42	2426
3VV	0.00	0.00	0.00	183
4CH	0.65	0.72	0.68	4210
LVOT	0.32	0.06	0.11	301
RVOT	0.00	0.00	0.00	92
Weighted Acc	0.53	0.57	0.54	7212

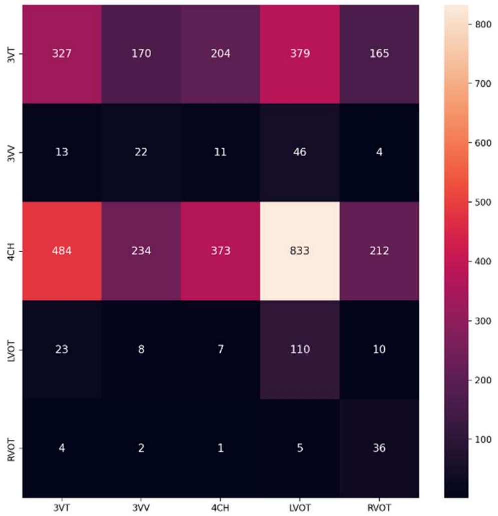

Fig. 8 Confusion Matrix after shuffling, striding, and stratified batching on the dataset.

Although the results are still bad, as the accuracy and loss don't even converge and the weighted average F1-score lowers to 0.27 in Table II, we can observe a reduction in bias towards most represented classes, as it can identify all the classes now and confounds itself more between them.

B. Augmentation

For the next improvement to the dataset, we can generate more instances of the less prevalent classes, so the model can see them more times during training. This is easily achieved with image augmentation techniques. Before going forth with our findings, we give a warning: any change to medical images should be evaluated by a professional in the area, as altering them too forcefully could result in unrealistic data being fed to the model and confounding the view classes features. We consulted a cardiologist that validated our proposed augmentation methods and other image manipulation techniques applied in our research. Not all of them will be presented here. For this paper, we chose only two: rotation and brightness.

Different from other works where augmentation is applied to all images of the training set, we use augmentation to increase the number of samples in the underrepresented classes 3VV, LVOT, and RVOT, as further explained below.

Rotation

As the name suggests, we transformed our images by rotating samples of the underrepresented classes. This was accomplished by using the scikit-image library [18], which is an open-source image processing library for the Python programming language that includes segmentation, geometric transformations, and many others.

We applied the rotate function, turning the images by a random angle in the range from -20° to +20°for each epoch. Rotating an exam image makes sense during training because the orientation of the heart is not standardized in the views, and what matters is the section plane.

Brightness

We decided to implement multiplicative brightness change in part due to its simplicity [18]. This process is defined by the following brightness and contrast adjustment equation:

$$T(i,j) = \alpha . I(i,j) + \beta \qquad (1)$$

where i, j are pixels coordinates, I is the input image, T is the transformed image, α represents gain, and β represents bias. The change in brightness can easily be done by just multiplying each pixel by a set value close to 1 and then clipping it into the range [0, 1].

Small changes to the exam's brightness help the network identify features affected by shadows and generalize equipment differences and/or tissue density differences.

Another round of training was executed with all proposed methods. As mentioned before, the rotation and brightness augmentation were only applied to the less favored classes, effectively creating new samples. In Fig. 9, we can see the increase of those samples on the training set. This increase allows us to choose larger batch sizes for the stratified batching. Using all these techniques helped reduce the difference between the most prevalent class (4CH) and the less prevalent class (RVOT) from 45 to around 11 times more frames than the other.

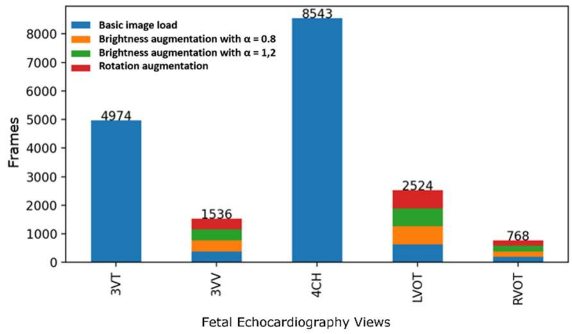

Fig. 9 Training set after preprocessing and augmentation methods. Blue are samples loaded normally, Orange are samples loaded with brightness augmentation with alpha = 0.8, Green are samples loaded with brightness augmentation with alpha = 1.2, and Red are samples loaded with rotation augmentation.

TABLE II.
SHUFFLE, STRIDE AND STRATIFIED BATCHING TEST REPORT

Class	Precision	Recall	F1-score	Number of Samples
3VT	0.38	0.26	0.31	1245
3VV	0.05	0.23	0.08	96
4CH	0.63	0.23	0.27	2136
LVOT	0.08	0.70	0.14	158
RVOT	0.08	0.75	0.15	48
Weighted Acc	0.50	0.24	0.27	3683

In Fig. 10 and Fig. 11, the learning history and confusion matrix can be further examined.

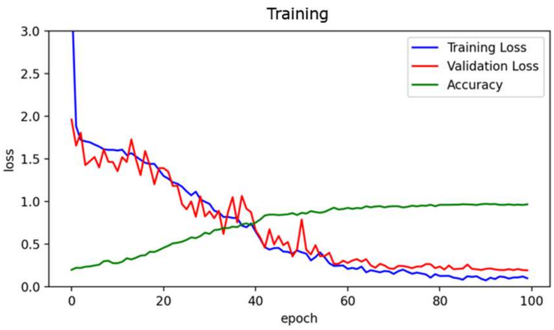

Fig. 10 Learning History after adding rotation and brightness augmentation

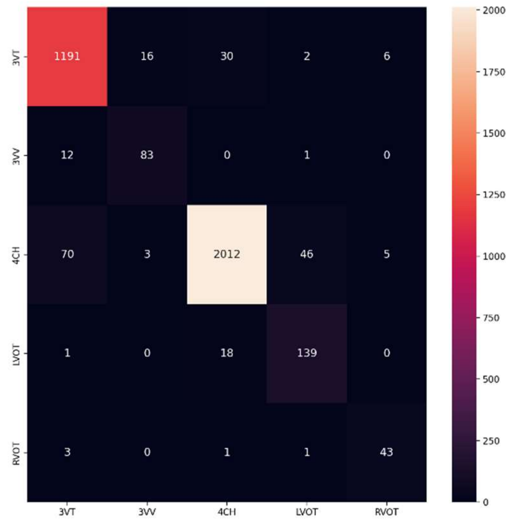

Fig. 11 Confusion Matrix after rotation and brightness augmentation

Finally, we can see accuracy and loss converging and a great leap in the model performance as seen in Table III. All the less prevalent classes have been learned by the model and the main view (4CH) has also increased to an astonishing F1-score of 0.94.

IV. Conclusions

Recapping, we have proposed in this paper that enhancing the quality of a small and unbalanced dataset by use of preprocessing and augmentation methods could reduce class imbalance and improve the model performance. Through many experiments and setbacks, we were able to analyze the impact of these methods on the dataset, observing the model go from a very poor classification to a remarkable F1-Score of 0.94 with a confusion matrix with few false positives and false negatives.

We can conclude that, if we take the time to understand better the data being used and apply the right methods for improvement of the dataset, like reducing redundancy, increasing diversity, and augmenting the images, it is possible to bring balance to a small, unbalanced dataset. Also, using a proven, efficient model is not enough to obtain good performance if your dataset is bad or too raw for use.

Exploration of the quality of the dataset has a lot of potentials, especially in medical imaging. Some promising subjects that we are already discussing and researching for future work are using transfer learning in known models trained with computer vision datasets, for feature extraction and fine-tuning of our model [19].

Another possibility is to explore a weighted loss function in conjunction with the methods present here, as this loss would punish the network more accordingly weight defined for each class [20]. We already have some crude implementations of this technique but was not used in this paper.

Also, many professionals in the medical field feel uneasy with the neural networks' black box nature and would give more credit for these solutions if they could understand how they have taken decisions. Explainable Artificial Intelligence (XAI) aims to solve these issues by bringing more transparency to the understanding of the model's bias, decisions impact and, as a result, generating better metrics [21].

References

[1] Luis Perez and Jason Wang. The Effectiveness of Data Augmentation in Image Classification using Deep Learning. *arXiv:1712.04621 [cs]*, December 2017. arXiv: 1712.04621.

[2] S. Kevin Zhou, Hayit Greenspan, Christos Davatzikos, James S. Duncan, Bram Van Ginneken, Anant Madabhushi, Jerry L. Prince, Daniel Rueckert, and Ronald M. Summers. A Review of Deep Learning in Medical Imaging: Imaging Traits, Technology Trends, Case Studies With Progress Highlights, and Future Promises. *Proceedings of the IEEE*, 109(5):820–838, May 2021, doi:10.1109/JPROC.2021.3054390.

[3] Andre Esteva, Katherine Chou, Serena Yeung, Nikhil Naik, Ali Madani, Ali Mottaghi, Yun Liu, Eric Topol, Jeff Dean, and Richard Socher. Deep learning-enabled medical computer vision. *npj Digital Medicine*, 4(1):5, December 2021, doi:10.1038/s41746-020-00376-2.

[4] Swati Rai, Jignesh S. Bhatt, and S. K. Patra. Augmented Noise Learning Framework for Enhancing Medical Image Denoising. *IEEE Access*, 9:117153–117168, 2021,doi:10.1109/ACCESS.2021.3106707.

Table III.
Rotation and Brightness Augmentation added Test Report

Class	Precision	Recall	F1-score	Number of Samples
3VT	0.93	0.96	0.94	1245
3VV	0.81	0.86	0.84	96
4CH	0.98	0.94	0.96	2136
LVOT	0.74	0.88	0.80	158
RVOT	0.80	0.90	0.84	48
Weighted Acc	0.94	0.94	0.94	3683

[5] Jing Wang, Xiaofeng Liu, Fangyun Wang, Lin Zheng, Fengqiao Gao, Hanwen Zhang, Xin Zhang, Wanqing Xie, and Binbin Wang. Automated interpretation of congenital heart disease from multi-view echocardiograms. *Medical Image Analysis*, 69:101942, April 2021, doi:10.1016/j.media.2020.101942.

[6] Amirata Ghorbani, David Ouyang, Abubakar Abid, Bryan He, Jonathan H. Chen, Robert A. Harrington, David H. Liang, Euan A. Ashley, and James Y. Zou. Deep learning interpretation of echocardiograms. *npj Digital Medicine*, 3(1):10, December 2020, doi:10.1038/s41746-019-0216-8.

[7] Day, T.G.; Kainz, B.; Hajnal, J.; Razavi, R.; Simpson, J.M. Artificial Intelligence, Fetal Echocardiography, and Congenital Heart Disease. *Prenatal Diagnosis* 2021, 41, 733–742, doi:10.1002/pd.5892.

[8] Nurmaini, S.; Rachmatullah, M.N.; Sapitri, A.I.; Darmawahyuni, A.; Tutuko, B.; Firdaus, F.; Partan, R.U.; Bernolian, N. Deep Learning-Based Computer-Aided Fetal Echocardiography: Application to Heart Standard View Segmentation for Congenital Heart Defects Detection. *Sensors* 2021, *21*, 8007, doi:10.3390/s21238007.

[9] Fiorentino, M.C.; Villani, F.P.; Di Cosmo, M.; Frontoni, E.; Moccia, S. A Review on Deep-Learning Algorithms for Fetal Ultrasound-Image Analysis. arXiv:2201.12260 [cs, eess] 2022.

[10] Nisselrooij, A.E.L.; Teunissen, A.K.K.; Clur, S.A.; Rozendaal, L.; Pajkrt, E.; Linskens, I.H.; Rammeloo, L.; Lith, J.M.M.; Blom, N.A.; Haak, M.C., "Why are congenital heart defects being missed?", Ultrasound Obstet Gynecol, vol. 55, nº 6, p. 747–757, jun. 2020, doi: 10.1002/uog.20358.

[11] Dablain, D.; Krawczyk, B.; Chawla, N.V. DeepSMOTE: Fusing Deep Learning and SMOTE for Imbalanced Data. *IEEE Transactions on Neural Networks and Learning Systems* 2022, 1–15, doi:10.1109/TNNLS.2021.3136503.

[12] Mullick, S.S.; Datta, S.; Das, S. Generative Adversarial Minority Oversampling. *In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*; October 2019.

[13] Bellinger, C.; Corizzo, R.; Japkowicz, N. Calibrated Resampling for Imbalanced and Long-Tails in Deep Learning. *In Proceedings of the Discovery Science*; Soares, C., Torgo, L., Eds.; Springer International Publishing: Cham, 2021; pp. 242–252, doi: 10.1007/978-3-030-88942-5_19.

[14] Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. *In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*; 2017; pp. 2999–3007, doi: 10.1109/ICCV.2017.324.

[15] Chollet, Francois. 2017. *Deep Learning with Python*. New York, NY: Manning Publications.

[16] Mingxing Tan and Quoc V. Le. "Efficientnet: Rethinking model scaling for convolutional neural networks", 2020.

[17] Vaseli, H.; Liao, Z.; Abdi, A.H.; Girgis, H.; Behnami, D.; Luong, C.; Taheri Dezaki, F.; Dhungel, N.; Rohling, R.; Gin, K.; et al. "Designing Lightweight Deep Learning Models for Echocardiography View Classification. In Proceedings of the Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling"; Fei, B., Linte, C.A., Eds.; SPIE: San Diego, United States, March 8 2019; p. 14.

[18] Scikit-image library documentation, https://scikit-image.org/docs/stable/

[19] Long Teng, Zhong, Liang Fu, Qian Ma, Yu Yao, Bing Zhang, Kai Zhu, and Ping Li. Interactive Echocardiography Translation Using Few-Shot GAN Transfer Learning. *Computational and Mathematical Methods in Medicine*, 2020:1–9, March 2020.

[20] K. R. M. Fernando and C. P. Tsokos, "Dynamically Weighted Balanced Loss: Class Imbalanced Learning and Confidence Calibration of Deep Neural Networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 7, pp. 2940–2951, 2022, doi: 10.1109/TNNLS.2020.3047335..

[21] Qiao, S.; Pang, S.; Luo, G.; Pan, S.; Yu, Z.; Chen, T.; Lv, Z. "RLDS: An Explainable Residual Learning Diagnosis System for Fetal Congenital Heart Disease". Future Generation Computer Systems 2022, 128, 205–218, doi:10.1016/j.future.2021.10.001.