Artificial Intelligence/based systems development and integrated solutions for the social and health protection network

Angelo Lasala[1,*], Sara Mazzucato[1,*], Andrea Bandini[2,3], Silvestro Micera[1,4] and Sara Moccia[1]

[1]*The BioRobotics Institute and the Department of Excellence in Robotics and Artificial Intelligence, Scuola Superiore Sant'Anna, Pisa, Italy, Viale Rinaldo Piaggio 34, 56025, Pontedera (PI), Italy*

[2]*Health Science Interdisciplinary Research Center, Scuola Superiore Sant'Anna, Pisa, Italy*

[3]*KITE Research Institute - Toronto Rehabilitation Institute - University Health Network, 550 University Avenue, Toronto, ON M5G 2A2, Canada*

[4]*Bertarelli Foundation Chair in Translational Neural Engineering, Center for Neuroprosthetics and Institute of Bioengineering, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland*

Abstract

Decentralizing part of specialist evaluations, some diagnostic procedures and some care activities becomes necessary for allowing citiziens to access healthcare services in decentralized cities as much as possible, and guarantee all citizens equal access to healthcare services. This contribution describes our research inside the Proximity Care Project, which aims at developing expert or artificial intelligence-based systems that support the correct identification of subjects at risk of pathology development. The activities here presented the design of (1) a predictive model for assisted diagnosis of pathologies from the analysis of electronic medical records; (2) ultrasound fetoscopy image analysis algorithms for automating the image acquisition of ultrasound fetal head standard planes and supporting the diagnosis. These findings could facilitate personalized therapy planning by predicting diseases risk from electronic medical records and also allows us to move toward model interpretability of image algorithms by looking at the discriminative elements in the ultrasound image.

Keywords

Artificial Intelligence, Electronic Health Records, Natural Language Processing, Fetal Head Standard Planes, Conditional Generative Model

1. Introduction

The Proximity Care project (https://www.santannapisa.it/en/health-science/proximity-care) proposes to implement a series of technological solutions that fit into a pathway involving multi-professional teams (general practitioners, specialists, community nurses, third sector operators, ...) and the patient.

The Health Profile of the Valle del Serchio district area for the year 2020 shows that the prevalence of chronicities is higher than the regional average. In particular, with regard to heart failure as visible in Fig. 1, the Valle del Serchio district zone is the one in Tuscany with the highest prevalence, for ischaemic heart disease it is in second place, while the prevalences for stroke and Chronic Obstructive Pulmonary Disease (COPD) are in the regional range. Hospitalisation rates in both males and females are generally higher than the average for the Azienda USL Toscana nord ovest. Additionally, although the downward trend in mortality in Tuscany has been consolidated for many years, the standardised mortality rate for males resident in the Serchio Valley was the worst in all of Tuscany and increased between 2013 and 2016. Looking at the provision of visits, a gap emerges between first visits and control visits: this gap is particularly evident for cardiology visits. When comparing the crude rates of provision of first visits and cardiology check-ups in 2020, it emerges that first visits are in line with the rates of the other district areas, while for check-ups there is a sharp decrease.

In order to allow all citizens an equal access to healthcare services, it becomes mandatory to decentralise part of the specialist evaluations, some diagnostic procedures and some care activities. This need emerged in a pressing manner during the pandemic, even in the geographical areas of Tuscany with the easiest access to healthcare services.

Our main role in the project is to develop artificial intelligence algorithms for the analysis of clinical data collected in dispersed locations in order to support the work of general practitioners and medical specialists. The work involves two main activities:

Ital-IA 2023: 3rd National Conference on Artificial Intelligence, organized by CINI, May 29–31, 2023, Pisa, Italy

*Shared first authorship

angelo.lasala@santannapisa.it (A. Lasala);
sara.mazzucato@santannapisa.it (S. Mazzucato);
andrea.bandini@santannapisa.it (A. Bandini);
silvestro.micera@santannapisa.it (S. Micera);
sara.moccia@santannapisa.it (S. Moccia)
0000-0002-5042-0834 (A. Lasala); 0000-0002-3509-2887 (A. Bandini); 0000-0003-4396-8217 (S. Micera); 0000-0002-4494-8907 (S. Moccia)

CEUR Workshop Proceedings (CEUR-WS.org)

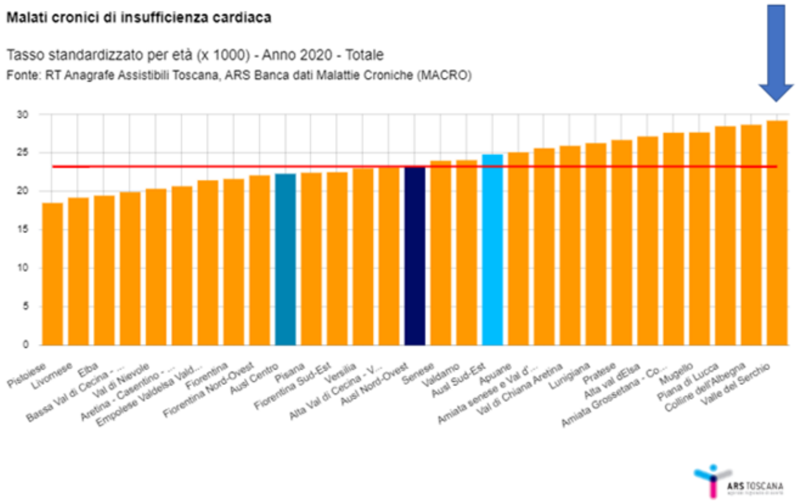

Figure 1: Tuscany heart failure cronic patiets in 2020

1. Analysis of data extracted from electronic medical records for the screening of diseases affecting the general population (e.g. cardiomyopathies, diabetes, etc.)
2. Analysis of ultrasound images to automate the image acquisition procedure (possibly by means of a robotic platform) and support screening activities

For the image analysis, the potential of a class of ML algorithms called deep-learning (DL) was investigated. To date, DL represents the state of the art for the analysis of ultrasound images. While waiting to collect a local dataset, datasets of ultrasound images available online were used.

2. Predictive models for amyloidosis screenig from EHR analysis

Amyloidosis (AM) is a group of diseases caused by abnormal protein folding that leads to the formation of insoluble fibrils in tissues and organs [1]. Its prognosis is poor, with cardiac death being the most common outcome [2]. In the US, there are 1275 to 3200 new cases per year, with an annual incidence ranging from 9.7 to 14.0 cases per million person-years. The total annual healthcare costs for patients range from $92,513$ to $114,030$ [3]. EHRs are an essential part of the clinical data generated worldwide, but the diagnosis of amyloidosis is often delayed [4].

This study aimed to develop an algorithm to screen patients' health status automatically by using EHR data and ML. The objectives were to identify patients with AM using clinically important information obtained from EHR and to foster model interpretability by analyzing the importance of risk factors. A case-control study was conducted to analyze the risk factors that determined the development of AM. For this purpose, before the predictive algorithms can be developed, it is necessary to apply Natural Language Processing (NLP) methods in order to be able to represent the unstructured text noted down by medical staff in the EHR in such a way that it can be processed by the machine-learning (ML) algorithms. So, the first step was the collection of clinical datasets in order to train the NLP algorithms and thus structure all the fields of the EHR.

Data from 418 patients with cardiac conditions were analyzed, recorded between 2010 and 2022 in Italy. Of these, 205 had AM and 213 had non-amyloid heart failure (HC). The study looked at 12 risk factors identified in previous research [5, 6, 7]. Missing values were replaced with null values and duplicates were removed. All participants gave informed consent and the study was conducted in accordance with ethical guidelines.

The study compared three supervised learning algorithms (Random Forest, Support Vector Machine, and XgBoost) to classify in Hc or AM group. These classifiers were chosen for their interpretability, accuracy, and low computation effort [8]. Random forests combine the predictions of several decision trees, each trained independently [9]. SVMs are commonly used when data is not linearly separable, mapping features to a higher-dimensional space using kernel functions [10, 11]. XgBoost is a tree-based boosting algorithm that optimizes

Table 1
Median (1° quartile, 3° quartile) of precision (P), recall (R) and f1-score (f1) obtained with the three classifiers

	RF	SVM	XGB
P	0.72 (0.68, 0.75)	0.64 (0.57, 0.68)	0.72 (0,69, 0.77)
R	0.76 (0.57, 0.81)	0.61 (0.54, 0.69)	0.78 (0.75, 0.80)
f1	0.73 (0.71, 0.75)	0.61 (0.58, 0.68)	0.74 (0.72, 0.77)

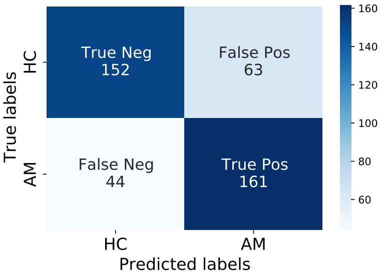

Figure 2: Confusion matrix obtained with the best RF.

new trees for every iteration, resulting in fewer false alarms and accurate classification [12]. Each algorithm was optimized with various hyper-parameters.

Hyperparameter tuning was performed for each classifier using a grid search to find the best parameters. Stratified 10-fold cross-validation was used to evaluate performance, and accuracy, precision, recall, and F1-measure were compared. The set of tested hyperparameters is reported.

1. RF: The number of trees in the forest, `n_estimations`:[50, 60, 70, 80, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600], the maximum depth of the tree, `max_depth`:[3,4,5,6,7].
2. SVM: `C`:[1, 10, 100, 1000], `kernel`:['linear']; `C`:[1, 10, 100, 1000], `kernel`:['rbf'], `gamma`:[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]; `C`:[1, 10, 100, 1000], `kernel`:['poly'], `degree`:[2,3,4], `gamma`:[0.01,0.02,0.03,0.04,0.05]
3. XGB: `max_depth`: range(2, 10, 1), `n_estimations`: range(60, 220, 40), `learning_rate`: [0.1, 0.01, 0.05]

Table 1, as well as Fig. 2 and Fig. 3, present the classification performance and confusion matrices obtained using the three algorithms.

This study aimed to evaluate three machine learning classifiers to identify the presence of amyloidosis using data from electronic health records. The Random Forest (RF) classifier had the highest median and lower interquartile range and provided insights on the importance of predictors for identifying amyloidosis. So, RF was identified as the best classifier, as indicated by its superior classification performance in Table 1, in accordance with other studies like [13]. To further investigate RF's performance, its feature importance were examined. The analysis revealed that the top five most important features, in descending order, were EF, NYHA, DY, DB, and HY. These results are presented in Fig. 3 and are consistent with previous literature [14]. EF was highlighted as a quantitative and objective parameter in prognostic models as in [15], while NYHA was significantly associated with the disease in patients with AL amyloidosis [16]. Dyslipidemia and diabetes are potential predictors that need further research. A cohort of patients with amyloidosis and diabetes had worse outcomes than those with atrial fibrillation, according to the study in [17].

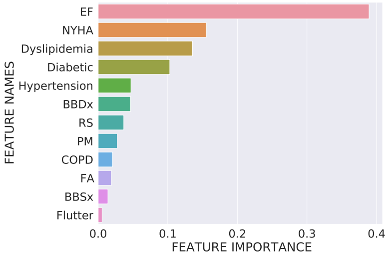

Figure 3: Feature importance of RF.

In conclusion, this study presented an automatic method for identifying patients at risk of AM using electronic medical records. Different classifiers were compared, and RF had the best performance. Feature importance results were consistent with major risk factors identified in literature, providing additional medical insights. These findings could facilitate personalized therapy planning by predicting amyloidosis risk from electronic medical records.

3. Medical image analysis for supporting clinicians

Due to its portability, low-cost and non-invasive nature, physicians and radiologists extensively use ultrasound (US) imaging for screening and diagnosis. However, US acquisition also presents unique challenges, such as low imaging quality caused by noise and artifacts, high dependence on operator experience, and high inter- and intra-variability across different patients and manufacturers' US systems [18]. Moreover, to the extent of achieving better reproducibility of biometric assessments [19], international organizations such as the International Society for Ultrasounds in Obstetrics and Gynaecology (ISUOG) and the American Society of Echocardiography Interna-

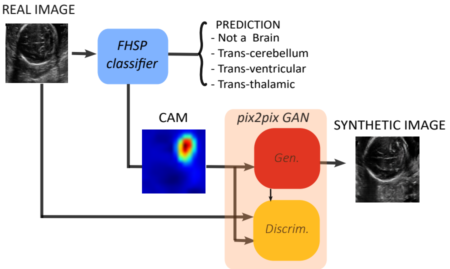

Figure 4: Workflow of proposed approach to exploit CAM in the generation of FHSP US images.

tional Alliance Partner provided worldwide guidelines for US standard plane acquisition in the related medical field [20, 21].

Thanks to their ability to tackle the challenges in US imaging, DL algorithms are used to support the analysis of the US acquisition [22], and to assist healthcare operators in the standard planes identification [19]. Nonetheless, two paramount issues arise when exploiting the DL algorithms in medical treatments: i) the outcomes of the model are referred to as black-boxes due to a lack of understanding of the decision-making process [23], and ii) the need for large annotated datasets to train DL algorithms is still perceived by the US-image analysis community as a major bottleneck to the development of robust algorithms [19]. Therefore, we propose a novel approach that aims to overcome these issues both in terms of model interpretability and model performance. In particular, our approach is based on two main steps:

1. Internal network representation: the visualization of the most discriminative elements (saliency or class activation maps) of the US image with the aim to understand the decision-making process of the model;
2. Imaging generation: synthesizing realistic US images exploiting the internal network representation to enlarge the training dataset.

We applied the proposed approach in the evaluation of fetal's growth during gestational screening. Obtaining fetal head standard planes (FHSPs) is of fundamental importance to visualize cerebral structures and diagnose neural anomalies during the mid-semesters of gestation. With the aim to improve the quality of these US acquisitions, we apply the proposed approach to the identification of the three common FHSPs: the trans-ventricular (TV), trans-thalamic (TT), and trans-cerebellar (TC) planes [24].

Fig. 4 illustrates the proposed approach to exploit the saliency maps as prior in the generation of synthetic images. Firstly, to obtain class activation cams (CAMs) [25] from FHSPs images, VGG16 [26] was used to classify TV, TT, and TC planes. CAMs were then computed as:

Table 2
Precision, recall, and f1-score for FHSPs classifications on the test set released in [24]

	precision	recall	f1-score	number of images
TC	0.86	0.86	0.86	339
TT	0.86	0.76	0.81	765
TV	0.66	0.84	0.74	302

$$CAM^c = \sum_{k=1}^{K} w_k^c A^k \tag{1}$$

where A^k is the $(14, 14)$ k-th feature of the last convolutional layer of VGG16 and w_k^c is the k-th weight corresponding to class c, with $c = \{\text{TV}, \text{TT}, \text{TC}\}$, and the total number of features $K = 512$. Once generated, the CAMs were successively used as the condition for conditional generative adversarial network (cGAN) to control the appearance of the generated samples. We selected the pix2pix cGAN [27] architecture, a cGAN that has been explored mainly in the clinical literature showing promising results [28]. Briefly, in conditional generative models the generator and the discriminator are trained to fool each other, in particular, the generator is forced to synthesize realistic images and the discriminator to identify as real the outcome of the generator.

The performances of the classification are reported in Table 2. VGG16 identifies better the TC and TT planes than the TV ones. To compare the results with the literature, we computed also the mean average accuracy, which is equal to 82.1 ± 5.2 %, in line with the human accuracy reported in [24].

Then, the CAMs are obtained by equation (1). As reported in Fig. 5, through CAM we can identify the discriminative element of each FHSP, namely the atrium for the TV (top left image), the thalamus for the TT (to right image), and the cerebellum for the TC (bottom image). Notably, with this approach, we can understand where the model looking at to make its prediction.

Finally, Fig. 6 shows some of the synthetic images (bottom row of each panel) generated from the saliency maps (top row of each panel). Two qualitative observations arise by comparing the synthetic images with the original ones (middle row of each panel). First, the position of the fetal head is not centered in the images, but shows variability across images according to the CAM. Second, the region surrounding the head contains details that remind real FHSPs images. Overall, the proposed model allows generating images that are similar to the outcome of the US-imaging acquisition. Although our results are promising, a limitation of this work is the demanding

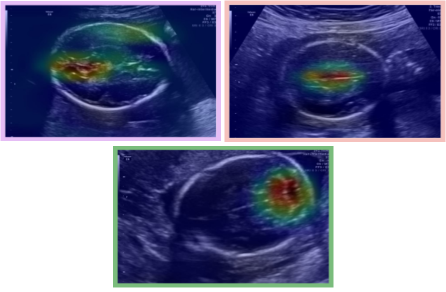

Figure 5: Class activation maps of FHSP: (top left image) TV, (top right image) TT, and (bottom image) TC.

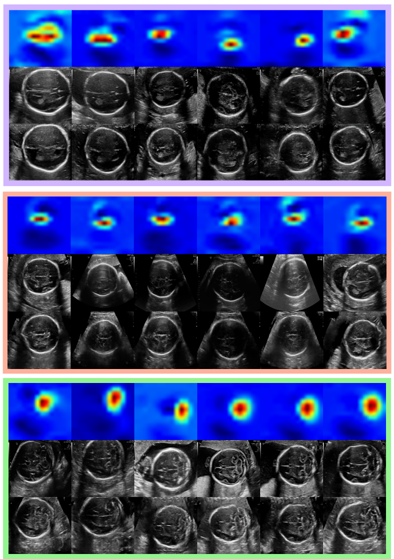

Figure 6: Generated images: (top panel) TV, (middle panel) TT, and (bottom panel) TC. The synthetic images (bottom row of the panels) are generated from the CAMs (top row of the panels) and compared with the real ones (middle row of the panels). Each panel contain 6 samples, arranged in row, correctly identified during classification.

generation of some characteristic landmarks, especially for TT. Precisely, these issues may derive from the weak ability of the proposed model to correctly identify this class, and the lack of details in training samples.

In conclusion, we want to underline that the proposed approach allows us to move toward model interpretability of the DL algorithm medical application by looking at the discriminative elements in the US image. Moreover, we are confident that exploiting CAMs to generate FHSPs using raw US images will be a valid augmentation tool to improve the performance of FHSPs classification problems in case of reduced datasets.

4. Conclusion

In this paper, we presented our preliminary results on our artificial intelligence-based systems inside the Proximity Care Project. They aim to assist in the screening of amyloidosis from the analysis of EHR and generate realistic FHSP ultrasound images in a real medical scenario. Despite the prominent results, using ML and DL algorithms limitations such as the paucity of the datasets and the interpretability of the models.

Future work will expand our dataset, conduct automatic feature selection, and use longitudinal data to develop ML tools for predicting the risk of amyloidosis for personalized therapies. Additionally, the development of NLP techniques using medical records from the *Fondazione Toscana Gabriele Monastiero* is in progress to create predictive algorithms for detecting amyloidosis. The study involves two steps: structuring medical records and applying predictive models to find correlations between patients' symptoms. The research will be extended to the medical records of general practitioners in the Garfagnana area. In the field of US medical imaging analysis, the effort to improve model interpretability must be a central role besides the performance of DL algorithms. In this regard, our approach shows that the class activations mapping represents a powerful tool to move forward model interpretability. Moreover, we exploit the CAM to synthesize new images to enlarge the dataset to further improve the performance of the DL model both in terms of accuracy and computational time. Therefore, the future step is to deploy our FHSPs classification model into an ultrasound probe. The goal is to return in real-time the outcomes of the model to assist the healthcare operator during the US image acquisition of standard planes.

5. Acknowledgment

The authors would like to acknowledge the funding support from Proximity Care Project aimed at, technological innovation for the social and health protection network of inland areas in the province of Lucca (Italy). A Project by Scuola Sant'Anna di Pisa - Interdisciplinary Center Health Science with the support of Fondazione Cassa di Risparmio di Lucca, in collaboration with: Azienda Usl Toscana Nord Ovest, Fondazione Monasterio, Unione dei Comuni della Garfagnana, Unione dei Comuni della Valle del Serchio, Conferenza dei Sindaci della Valle del Serchio.

References

[1] J. W. Kelly, Mechanisms of amyloidogenesis, Nature Structural Biology 7 (2000) 824–826.

[2] S. Yamada, A. Yoshihisa, N. Hijioka, M. Kamioka, T. Kaneshiro, T. Yokokawa, T. Misaka, T. Ishida, Y. Takeishi, Autonomic dysfunction in cardiac amyloidosis assessed by heart rate variability and heart rate turbulence, Annals of Noninvasive Electrocardiology 25 (2020) e12749.

[3] T. P. Quock, T. Yan, E. Chang, S. Guthrie, M. S. Broder, Epidemiology of al amyloidosis: a real-world study using us claims data, Blood Advances 2 (2018) 1046–1053.

[4] J. S. Cavalier, R. Loungani, M. G. Khouri, Red flag comorbid conditions that are easily extracted from the electronic health record often precede diagnosis of cardiac amyloidosis by years, Circulation 144 (2021) A13585–A13585.

[5] G. Vergaro, A. Aimo, A. Barison, D. Genovesi, G. Buda, C. Passino, M. Emdin, Keys to early diagnosis of cardiac amyloidosis: red flags from clinical, laboratory and imaging findings, European Journal of Preventive Cardiology 27 (2020) 1806–1815.

[6] G. Vergaro, A. Aimo, C. Rapezzi, V. Castiglione, I. Fabiani, A. Pucci, G. Buda, C. Passino, J. Lupón, A. Bayes-Genis, et al., Atrial amyloidosis: mechanisms and clinical manifestations, European Journal of Heart Failure (2022).

[7] V. Castiglione, M. Franzini, A. Aimo, A. Carecci, C. M. Lombardi, C. Passino, C. Rapezzi, M. Emdin, G. Vergaro, Use of biomarkers to diagnose and manage cardiac amyloidosis, European Journal of Heart Failure 23 (2021) 217–230.

[8] R. Rosati, L. Romeo, G. Cecchini, F. Tonetto, P. Viti, A. Mancini, E. Frontoni, From knowledge-based to big data analytic model: a novel iot and machine learning based decision support system for predictive maintenance in industry 4.0, Journal of Intelligent Manufacturing 34 (2023) 107–121.

[9] M. Skurichina, R. P. Duin, Bagging, boosting and the random subspace method for linear classifiers, Pattern Analysis & Applications 5 (2002) 121–135.

[10] M. Bernardini, L. Romeo, P. Misericordia, E. Frontoni, Discovering the type 2 diabetes in electronic health records using the sparse balanced support vector machine, IEEE Journal of Biomedical and Health Informatics 24 (2019) 235–246.

[11] B. Schölkopf, A. J. Smola, F. Bach, et al., Learning with kernels: support vector machines, regularization, optimization, and beyond, MIT press, 2002.

[12] S. S. Dhaliwal, A.-A. Nahid, R. Abbas, Effective intrusion detection system using xgboost, Information 9 (2018) 149.

[13] M. Aria, C. Cuccurullo, A. Gnasso, A comparison among interpretative proposals for random forests, Machine Learning with Applications 6 (2021) 100094.

[14] L. Bonnefous, M. Kharoubi, M. Bézard, S. Oghina, F. Le Bras, E. Poullot, V. Molinier-Frenkel, P. Fanen, J.-F. Deux, V. Audard, et al., Assessing cardiac amyloidosis subtypes by unsupervised phenotype clustering analysis, Journal of the American College of Cardiology 78 (2021) 2177–2192.

[15] A. M. Maceira, S. K. Prasad, P. N. Hawkins, M. Roughton, D. J. Pennell, Cardiovascular magnetic resonance and prognosis in cardiac amyloidosis, Journal of Cardiovascular Magnetic Resonance 10 (2008) 1–11.

[16] A. V. Kristen, J. B. Perz, S. O. Schonland, U. Hegenbart, P. A. Schnabel, J. H. Kristen, H. Goldschmidt, H. A. Katus, T. J. Dengler, Non-invasive predictors of survival in cardiac amyloidosis, European journal of Heart Failure 9 (2007) 617–624.

[17] B. W. Sperry, M. N. Vranian, R. Hachamovitch, H. Joshi, A. Ikram, D. Phelan, M. Hanna, Subtype-specific interactions and prognosis in cardiac amyloidosis, Journal of the American Heart Association 5 (2016) e002877.

[18] S. Liu, Y. Wang, X. Yang, B. Lei, L. Liu, S. X. Li, D. Ni, T. Wang, Deep learning in medical ultrasound analysis: a review, Engineering 5 (2019) 261–275.

[19] M. C. Fiorentino, F. P. Villani, M. Di Cosmo, E. Frontoni, S. Moccia, A review on deep-learning algorithms for fetal ultrasound-image analysis, Medical Image Analysis 83 (2023) 102629.

[20] C. Mitchell, P. S. Rahko, L. A. Blauwet, B. Canaday, J. A. Finstuen, M. C. Foster, K. Horton, K. O. Ogunyankin, R. A. Palma, E. J. Velazquez, Guidelines for performing a comprehensive transthoracic echocardiographic examination in adults: recommendations from the american society of echocardiography, Journal of the American Society of Echocardiography 32 (2019) 1–64.

[21] L. Salomon, Z. Alfirevic, F. Da Silva Costa, R. L. Deter, F. Figueras, T. Ghi, P. Glanc, A. Khalil, W. Lee, R. Napolitano, A. Papageorghiou, A. Sotiriadis, J. Stirnemann, A. Toi, G. Yeo, ISUOG practice guidelines: ultrasound assessment of fetal biometry and growth, Ultrasound in Obstetrics & Gynecology 53 (2019) 715–723.

[22] Z. Lin, S. Li, D. Ni, Y. Liao, H. Wen, J. Du, S. Chen, T. Wang, B. Lei, Multi-task learning for quality assessment of fetal head ultrasound images, Medical Image Analysis 58 (2019) 101548.

[23] Z. Salahuddin, H. C. Woodruff, A. Chatterjee, P. Lambin, Transparency of deep neural networks for medical image analysis: A review of interpretability methods, Computers in Biology and Medicine 140 (2022) 105111.

[24] X. P. Burgos-Artizzu, D. Coronado-Gutiérrez, B. Valenzuela-Alcaraz, E. Bonet-Carne, E. Eixarch, F. Crispi, E. Gratacós, Evaluation of deep convolutional neural networks for automatic classification of common maternal fetal ultrasound planes, Scientific Reports 10 (2020) 1–12.

[25] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization, in: IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2921–2929.

[26] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556 (2014).

[27] P. Isola, J.-Y. Zhu, T. Zhou, A. A. Efros, Image-to-image translation with conditional adversarial networks, in: IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1125–1134.

[28] N. K. Singh, K. Raza, Medical image generation using generative adversarial networks: A review, Health informatics: A Computational Perspective in Healthcare (2021) 77–96.