





Automatic Detection of Extra-Cardiac Findings in Cardiovascular Magnetic Resonance

Dewmini Hasara Wickremasinghe¹(✉)() , Natallia Khenkina^{3,4}() ,
Pier Giorgio Masci^{1,2}() , Andrew P. King¹() , and Esther Puyol-Antón¹()

¹ School of Biomedical Engineering and Imaging Sciences, King's College London, London, UK

`dewmini.wickremasinghe@kcl.ac.uk`

² Guy's and St Thomas' Hospital, London, UK

³ Luigi Sacco University Hospital, Milan, Italy

⁴ Postgraduate School of Diagnostic and Interventional Radiology, Università degli Studi di Milano, Milan, Italy

Abstract. Cardiovascular magnetic resonance (CMR) is an established, non-invasive technique to comprehensively assess cardiovascular structure and function in a variety of acquired and inherited cardiac conditions. In addition to the heart, a typical CMR examination will also image adjacent thoracic and abdominal structures. Consequently, findings incidental to the cardiac examination may be encountered, some of which may be clinically relevant. We compare two deep learning architectures to automatically detect extra cardiac findings (ECFs) in the HASTE sequence of a CMR acquisition. The first one consists of a binary classification network that detects the presence of ECFs and the second one is a multi-label classification network that detects and classifies the type of ECF. We validated the two models on a cohort of 236 subjects, corresponding to 5610 slices, where 746 ECFs were found. Results show that the proposed methods have promising balanced accuracy and sensitivity and high specificity.

Keywords: Deep learning classification · Extra cardiac findings · Cardiovascular magnetic resonance

1 Introduction

Cardiovascular Magnetic Resonance (CMR) is a valuable, non-invasive, diagnostic option for the evaluation of cardiovascular diseases as it allows the assessment of both cardiac structure and function [21]. CMR is an advantageous image modality due to its wide field of view, accuracy, reproducibility and ability to scan in different planes. Also, due to its lack of exposure to ionizing radiation,

A. P. King and E. Puyol-Antón—Joint last authors.

© Springer Nature Switzerland AG 2021

B. W. Papież et al. (Eds.): MIUA 2021, LNCS 12722, pp. 98–107, 2021.

https://doi.org/10.1007/978-3-030-80432-9_8

CMR can be widely employed, except for patients with implanted electronic devices such as pacemakers [9, 15, 16]. The importance of the CMR imaging modality is evident from the fact that in the time period ranging from 2008 to 2018 CMR use increased by 573%, growing by the 14.7% in the year 2017–2018 alone [10].

A typical CMR acquisition begins with a gradient echo ‘scout’ image of several slices in the coronal, sagittal, and axial planes, followed by axial imaging of the entire chest, conventionally using a Half-Fourier Acquisition Single-shot Turbo spin Echo (HASTE) sequence [8]. The acquired images contain the heart, as well as significant portions of the upper abdomen and thorax. These regions may present extra-cardiac irregularities, which are defined as incidental Extra Cardiac Findings (ECFs) [8].

The importance of investigating the presence of incidental ECFs in CMR scans has been shown in previous studies [2, 8, 17, 22], and acknowledged by the European Association of Cardiovascular Imaging, which includes it as part of the European CMR certification exam [14]. In particular, Dunet *et al.* [2] reported a pooled prevalence of 35% of ECFs in patients undergoing CMR, 12% of which could be classified as major findings, i.e. requiring further investigation. ECFs are important for the early diagnosis of unknown diseases but can also be useful to determine the primary cardiac pathology which is being examined, due to the fact that some cardiac conditions have a multi-systemic environment [17]. In addition, when an ECF is identified, the important clinical question is whether the abnormality represents a benign or malignant lesion [17]. Two key examples are breast and lung cancer, since mammary and pulmonary tissue can be visualized on axial cross sectional imaging at the time of CMR. Previous works have shown that incidental breast lesions are identified in 0.1–2.5% of CMR studies and over 50% of these lesions are clinically significant [14–16, 52,]; similarly, the incidence of significant pulmonary abnormalities found in CMR examinations are up to 21.8% [2]. Another important factor to consider is that, depending on the institution, CMR examinations may be reported by cardiologists, radiologists or a combination thereof. A recent study showed that the highest accuracy to assess prevalence and significance of ECF in clinical routine CMR studies was reported when cardiologist and radiologist were working together [3], but this is not possible at all institutions. We believe that a computer-aided ECF detection tool could be beneficial in a clinical setting, especially when reporting is performed by only one specialist or by inexperienced operators.

The application of artificial intelligence (AI) in healthcare has great potential, for example by automating labour-intensive activities or by supporting clinicians in the decision-making process [19]. Liu *et al.* [12] compared the performance of clinicians to deep learning models in disease detecting tasks. They showed that AI algorithms performed equivalently to health-care professionals in classifying diseases from medical imaging. Although there is room for improvement, these results confirm the positive impact that AI could have in healthcare. Deep learning models have been employed in the automated detection of incidental findings in computed tomography (CT) and obtained promising results [1, 18, 23].

An automated pipeline able to detect the presence of incidental ECFs in CMR would not only be beneficial to the investigation of primary conditions and possible unknown diseases of the patient but could also reduce burden on overworked clinicians. In this paper, we investigate the feasibility of using deep learning techniques for the automatic detection of ECFs from the HASTE sequence.

2 Materials

This is a retrospective multi-vendor study approved by the institutional ethics committee and all patients gave written informed consent. A cohort of 236 patients (53.7 ± 15.7 years, 44% female) who underwent clinical CMR was manually reviewed to specifically assess the prevalence and importance of incidental ECFs. CMR image acquisitions were acquired with scanners of different magnetic field strengths and from different vendors with the following distributions: 70 subjects with 1.5T Siemens, 86 subjects with 1.5T Phillips and 80 subjects with 3.0T Phillips. From the CMR acquisitions, the HASTE sequence was used to detect any abnormal finding located outside the pericardial borders and the great vessels (aortic and pulmonary). ECFs were classified by anatomical location (i.e. neck, lung, mediastinum, liver, kidney, abdomen, soft tissue and bone) and by severity (i.e. major for findings that warrant a further investigation, new treatment, or a follow up e.g. lymphadenopathy or lung abnormalities; minor for findings that are considered benign conditions and don't require further investigations, follow up or treatment) [2, 13]. Of the 236 studies analysed, that correspond to 5610 slices, 746 ECFs were found. The distribution of ECFs by location and severity is shown in Fig. 1.

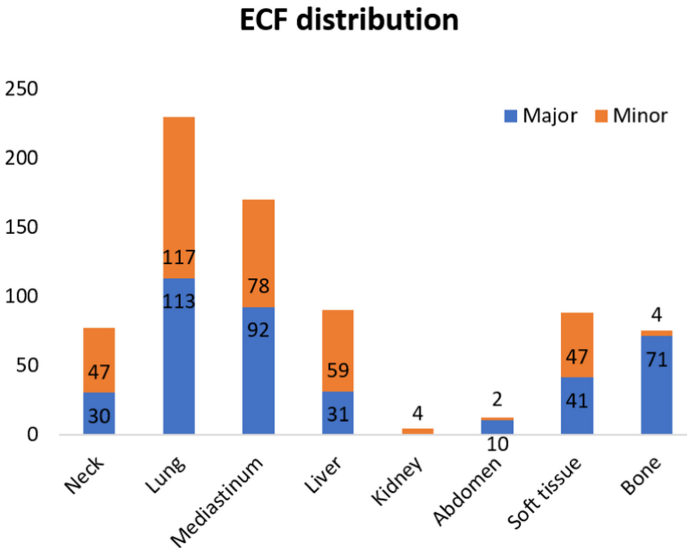


Fig. 1. Distribution of the ECFs split by location and severity.

3 Methods

The proposed framework for automatic detection of ECFs from the CMR HASTE images is summarized in Fig. 2, and each step is described below.

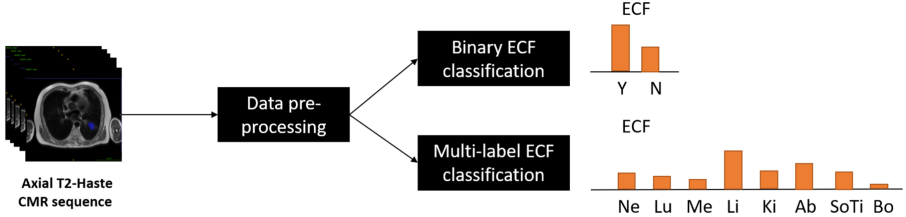


Fig. 2. Overview of the proposed framework for automatic ECF detection/classification.

3.1 Data Pre-processing

To correct for variation in acquisition protocols between vendors, all images were first resampled to an in-plane voxel size of 1.25×1.25 mm and cropped to a standard size of 256×256 pixels. The cropping was based on the centre of the images and the standard size was selected as the median size of all the images in the database. All DICOM slices were converted to numpy arrays and the pixel values were normalised between 0 and 1.

3.2 Binary ECF Classification

The first strategy aims to detect if any of the slices of the HASTE sequence has an ECF and we frame the problem as a binary classification task. We trained and evaluated seven state-of-the-art convolutional neural network (CNN) architectures: AlexNet [11], DenseNet [6], MobileNet [5], ResNet [4], ShuffleNet [24], SqueezeNet [7] and VGG [20].

3.3 Multi-label ECF Classification

The second strategy aims to not only detect the presence of ECFs but also identify to which class the ECF belongs. The chosen classes represent eight different areas of the body, namely neck, lung, mediastinum, liver, kidney, abdomen, soft tissue and bone. In this paper, we focus on the identification of ECFs and their subsequent classification based on the classes mentioned above, rather than their major/minor classification. As each slice can contain more than one ECF, we have framed the problem as a multi-label classification, which means that the output of the deep learning classifier supports multiple mutually non-exclusive classes. We extended the previous seven state-of-the-art CNN architectures to multi-label classification by using the number of classes as the number of nodes in the output layer and adding a sigmoid activation for each node in the output layer.

3.4 Training

The manually classified data were divided as follows: 80% were used for training and validation of the classification networks and 20% were used for testing. Data were split at patient level and bounded to a specific set, either training, validation or testing, in order to maintain data independence. Each network was trained for 200 epochs with binary cross entropy with a logit loss function. During training, data augmentation was performed on-the-fly using random translations (± 30 pixels), rotations ($\pm 90^\circ$), flips (50% probability) and scalings (up to 20%) to each mini-batch of images before feeding them to the network. The probability of augmentation for each of the parameters was 50%. Additionally, we implemented an adaptive learning rate scheduler, which decreases the learning rate by a constant factor of 0.1 after every 5 epochs, stopping at a plateau on the validation set (commonly known as ReduceLROnPlateau). This step was added as it improves training when presented with unbalanced datasets.

3.5 Statistics

The performance of the models was evaluated using a receiver operating characteristic (ROC) curve analysis, and based on this the balanced accuracy (BACC), sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were computed for the optimal classifier selected using the weighted Youden index. Sensitivity, also known as the true positive rate, is defined as the proportion of ground truth positively labelled examples that are identified as positive by the model; specificity, also known as true negative rate, is defined as the proportion of ground truth negatives that are identified as negative; PPV is defined as the proportion of identified positives that have a ground truth positive label; NPV is defined as the proportion of identified negatives with a ground truth negative label. For the multi-label classification algorithm, we extended this analysis to two conventional methods, namely micro-averaging and macro-averaging. Micro-averaging calculates metrics globally by counting the total true positives (TPs), true negatives (TNs), false positives (FPs) and false negatives (FNs), while macro-averaging calculates metrics for each label and finds their unweighted mean.

4 Results

4.1 Binary ECF Classification

Table 1 summarises the statistics computed from the results of the binary classification for each of the employed state-of-the-art networks.

Table 1. Mean sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and balanced accuracy (BACC) for the different binary ECF classifiers. Bold font highlights the best results.

	ALEX NET	DENSE NET	MOBILE NET	RES NET	SHUFFLE NET	SQUEEZE NET	VGG
Sensitivity	0.34	0.36	0.17	0.44	0.45	0.00	0.55
Specificity	0.80	0.77	0.93	0.79	0.72	1.00	0.76
PPV	0.19	0.18	0.25	0.23	0.18	NaN	0.24
NPV	0.90	0.90	0.89	0.91	0.91	0.88	0.93
BACC	0.57	0.57	0.55	0.62	0.59	0.50	0.66

The computed sensitivity values range between 0.34 and 0.55, except for MobileNet and SqueezeNet that obtained 0.17 and 0.00. Specificity and NPV, which are respectively the proportion of the correctly identified negative labels and the chance of the assigned label to be correct if identified as negative, obtained results close to 1. PPV values range between 0.18 and 0.25, although SqueezeNet, the only outlier, had NaN. BACC values fluctuate around 0.55. SqueezeNet obtained the lowest value (0.50) and VGG obtained 0.66, which is the highest computed BACC. It is noticeable that SqueezeNet performed poorly compared to the other networks and achieved the lowest computed values. On the other hand, the best performing network was VGG, which obtained the best sensitivity and BACC values.

4.2 Multi-label ECF Classification

Table 2 summarises the statistics computed from the results of the multi-label classification for each of the employed state-of-the-art networks.

As stated before, micro-averaging computes the metrics by calculating the total numbers of TPs, TNs, FPs and FNs. The best sensitivity value of 0.62 was obtained with AlexNet. Specificity and NPV, similarly to the binary classification task, have high values, often close to 1. The only outliers are AlexNet and ShuffleNet, which had specificity respectively 0.85 and 0.83. PPV had low results for all the networks and obtained the lowest values for AlexNet and ShuffleNet, which got respectively 0.21 and 0.20. It is noteworthy that although these networks obtained high values for the other metrics, they obtained the lowest values for PPV. The computed BACC values, in parallel to the ones computed in the binary case, oscillate between 0.50 and 0.74. The best BACC values were computed from AlexNet and ShuffleNet.

Macro-averaging calculates metrics per label and then finds their unweighted mean. In this case, sensitivity values are lower than the ones mentioned above. For macro averaging they range between 0.28 and 0.51, except for MobileNet and SqueezeNet that obtained the lowest sensitivities, respectively 0.09 and 0.00. As before, specificity and NPV values are close to the maximum. PPV obtained low results, between 0.13 and 0.34, except for SqueezeNet that obtained NaN. The NaN values are caused when the network predicts all the cases as negatives

Table 2. Mean sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and balanced accuracy (BACC) for the different multi-label ECF classifiers. Micro-averaging calculates metrics globally by counting the total true positives, true negatives, false positives and false negatives, while macro-averaging calculates metrics for each label and finds their unweighted mean. Bold font highlights the best results.

	ALEX NET	DENSE NET	MOBILE NET	RES NET	SHUFFLE NET	SQUEEZE NET	VGG
Micro sensitivity	0.62	0.44	0.12	0.40	0.57	0.00	0.32
Micro specificity	0.85	0.94	0.99	0.91	0.83	1.00	0.94
Micro PPV	0.21	0.32	0.41	0.23	0.20	NaN	0.28
Micro NPV	0.97	0.96	0.94	0.96	0.96	0.94	0.95
Micro BACC	0.74	0.69	0.56	0.65	0.70	0.50	0.63
Macro sensitivity	0.41	0.28	0.09	0.31	0.51	0.00	0.32
Macro specificity	0.84	0.93	0.99	0.91	0.83	1.00	0.94
Macro PPV	0.13	0.21	0.34	0.18	0.19	NaN	0.14
Macro NPV	0.97	0.96	0.94	0.95	0.96	0.96	0.95
Macro BACC	0.62	0.60	0.54	0.61	0.67	0.50	0.63

and therefore there are no true positives or false positives. This reflects the poor performance of the SqueezeNet network. The highest computed BACC was computed for ShuffleNet, while the other networks obtained values around 0.60. Again, MobileNet and SqueezeNet obtained the lowest values (0.54 and 0.50).

Visual results from the multi-label ECF classifier are shown in Fig. 3. The top row shows five images containing different ECFs (i.e. location and severity) which were correctly classified and the bottom row shows five images that have been misclassified by the network. Overall, it is apparent that the size and shape of the ECF can significantly vary. The performance of the network seems to be strongly influenced by the size of the ECF as well as the number of cases of that class of ECF in the training database.

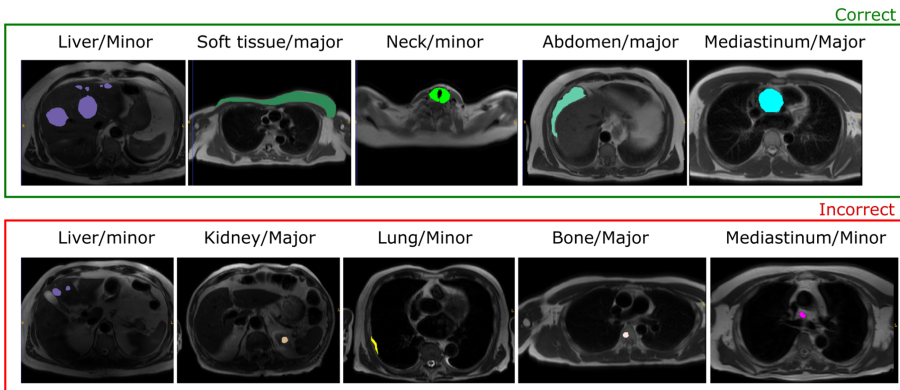


Fig. 3. Example results for proposed multi-label ECF classifier: top row shows correct cases and bottom row shows cases that have been misclassified.

5 Discussion and Conclusion

In a CMR examination, a careful assessment of non-cardiac structures may also detect relevant non-cardiac diseases. During a CMR acquisition, the inferior neck, entire thorax and upper abdomen are routinely imaged, particularly in the initial multi-slice axial and coronal images. Correctly identifying and reporting ECFs is beneficial to the patient and can prevent unnecessary over-investigation whilst ensuring that indeterminate or potentially important lesions are investigated appropriately. Cardiovascular diseases also often have systemic effects and the identification of ECFs can help with the interpretation of the primary cardiac pathology. In this paper, we have proposed for the first time a deep learning-based framework for the detection of ECFs from the HASTE sequence.

We approached the problem following two strategies: the first one consisted of a binary classification task that aimed to identify the presence of ECFs in each slice of the HASTE sequence; the second one consisted of a multi-label classification task which, in addition to the identification of ECFs, aimed to classify the ECFs based on their location.

For the first approach results showed that the best performing network was VGG, with BACC, sensitivity and specificity respectively equal to 66%, 55% and 76%. For the second approach the best performing networks were AlexNet and ShuffleNet, with a micro BACC higher than 70%, micro specificity above 83% and micro sensitivity values respectively of 62% and 57%. Macro metrics show that when computing the unweighted mean obtained from each label, performance decreases.

Our vision is that deep learning models could be used in clinical workflows to automate the identification of ECFs from CMR exams, thus reducing clinical workloads. This would require a high sensitivity to ensure that potential ECFs are not missed (i.e. minimise false negatives). False positives are less important as they can be eliminated by a subsequent cardiologist review. Therefore, performance is not currently sufficient for clinical needs. We believe that the main reasons for the low sensitivity are the limited amount of training data, the variation in image appearance due to the multi-vendor nature of the study and the large variation in the size, appearance and position of ECFs, as shown in Fig. 3. Obtaining more data, with less class imbalance would likely improve performance in future work. However, this work represents the first AI-based framework for automated detection and localization of ECFs in CMR images and therefore serves as a proof-of-principle. In future work, we will aim to gather more training data and develop novel techniques to improve the sensitivity of our models.

A limitation of the current framework is that we do not differentiate between major and minor ECFs and this is important in clinical practice to decide which ECF should be treated and which could be considered benign. We plan to address this in future work. We will also aim to combine a classification network with a segmentation network to allow localisation and differentiation of the different ECFs.

In conclusion, we have demonstrated the feasibility of using deep learning for the automatic screening of HASTE images for identifying potential ECFs.

Further work is required to improve the sensitivity of the technique and fully evaluate its role and utility in clinical workflows.

Acknowledgements. This work was supported by the EPSRC (EP/R005516/1 and EP/P001009/1) and the Wellcome EPSRC Centre for Medical Engineering at the School of Biomedical Engineering and Imaging Sciences, King's College London (WT 203148/Z/16/Z).

References

1. Blau, N., Klang, E., Kiryati, N., Amitai, M., Portnoy, O., Mayer, A.: Fully automatic detection of renal cysts in abdominal CT scans. *Int. J. Comput. Assist. Radiol. Surg.* **13**(7), 957–966 (2018)
2. Dunet, V., Schwitter, J., Meuli, R., Beigelman-Aubry, C.: Incidental extracardiac findings on cardiac MR: systematic review and meta-analysis. *J. Magn. Resonan. Imaging* **43**(4), 929–939 (2016)
3. Greulich, S., et al.: Extra cardiac findings in cardiovascular MR: why cardiologists and radiologists should read together. *Int. J. Cardiovasc. Imaging* **30**(3), 609–617 (2014)
4. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778 (2016)
5. Howard, A.G., et al.: MobileNets: efficient convolutional neural networks for mobile vision applications. arXiv preprint [arXiv:1704.04861](https://arxiv.org/abs/1704.04861) (2017)
6. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700–4708 (2017)
7. Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K., Dally, W.J., Keutzer, K.: SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size. arXiv preprint [arXiv:1602.07360](https://arxiv.org/abs/1602.07360) (2016)
8. Irwin, R.B., et al.: Incidental extra-cardiac findings on clinical CMR. *Eur. Heart J.-Cardiovasc. Imaging* **14**(2), 158–166 (2013)
9. Karamitsos, T.D., Francis, J.M., Myerson, S., Selvanayagam, J.B., Neubauer, S.: The role of cardiovascular magnetic resonance imaging in heart failure. *J. Am. Coll. Cardiol.* **54**(15), 1407–1424 (2009)
10. Keenan, N., et al.: UK national and regional trends in cardiovascular magnetic resonance usage—the British society of CMR survey results. *Eur. Heart J.* **41**(Supplement_2), ehaa946-0200 (2020)
11. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **25**, 1097–1105 (2012)
12. Liu, X., et al.: A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit. Health* **1**(6), e271–e297 (2019)
13. Mantini, C., et al.: Prevalence and clinical relevance of extracardiac findings in cardiovascular magnetic resonance imaging. *J. Thoracic Imaging* **34**(1), 48–55 (2019)
14. Petersen, S.E., et al.: Update of the European association of cardiovascular imaging (EACVI) core syllabus for the European cardiovascular magnetic resonance certification exam. *Eur. Heart J.-Cardiovasc. Imaging* **15**(7), 728–729 (2014)

15. Peterzan, M.A., Rider, O.J., Anderson, L.J.: The role of cardiovascular magnetic resonance imaging in heart failure. *Cardiac Failure Rev.* **2**(2), 115 (2016)
16. Rajiah, P., Desai, M.Y.: Cardiac magnetic resonance imaging-role and applications. *Eur. Cardiol.* **8**(1), 17–22 (2012)
17. Rodrigues, J.C., et al.: Extra-cardiac findings in cardiovascular magnetic resonance: what the imaging cardiologist needs to know. *J. Cardiovasc. Magn. Resonan.* **18**(1), 1–21 (2016)
18. Shenkman, Y., et al.: Automatic detection and diagnosis of sacroiliitis in CT scans as incidental findings. *Med. Image Anal.* **57**, 165–175 (2019)
19. Siegersma, K., Leiner, T., Chew, D., Appelman, Y., Hofstra, L., Verjans, J.: Artificial intelligence in cardiovascular imaging: state of the art and implications for the imaging cardiologist. *Netherlands Heart J.* **27**(9), 403–413 (2019)
20. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556) (2014)
21. Stacey, R.B., Hundley, W.G.: The role of cardiovascular magnetic resonance (CMR) and computed tomography (CCT) in facilitating heart failure management. *Curr. Treat. Options Cardiovasc. Med.* **15**(4), 373–386 (2013)
22. Wyttenbach, R., Médioni, N., Santini, P., Vock, P., Szucs-Farkas, Z.: Extracardiac findings detected by cardiac magnetic resonance imaging. *Eur. Radiol.* **22**(6), 1295–1302 (2012)
23. Yan, K., Wang, X., Lu, L., Summers, R.M.: DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. *J. Med. Imaging* **5**(3), 036501 (2018)
24. Zhang, X., Zhou, X., Lin, M., Sun, J.: ShuffleNet: an extremely efficient convolutional neural network for mobile devices. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6848–6856 (2018)