



Article

Automated Classification of Left Ventricular Hypertrophy on Cardiac MRI

Adam Budai ^{1,*}, Ferenc Imre Suhai ², Kristof Csorba ¹, Zsofia Dohy ², Liliana Szabo ², Bela Merkely ² and Hajnalka Vago ²

¹ Department of Automation and Applied Informatics, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics, H-1111 Budapest, Hungary; kristof@aut.bme.hu

² Heart and Vascular Center, Semmelweis University, H-1122 Budapest, Hungary; suhai.ferenc.imre@gmail.com (F.I.S.); dohyzsofi@gmail.com (Z.D.); sz.liliana.e@gmail.com (L.S.); merkely.bela@gmail.com (B.M.); vagoaha@gmail.com (H.V.)

* Correspondence: budai.adam@aut.bme.hu

Abstract: Left ventricular hypertrophy is an independent predictor of coronary artery disease, stroke, and heart failure. Our aim was to detect LVH cardiac magnetic resonance (CMR) scans with automatic methods. We developed an ensemble model based on a three-dimensional version of ResNet. The input of the network included short-axis and long-axis images. We also introduced a standardization methodology to unify the input images for noise reduction. The output of the network is the decision whether the patient has hypertrophy or not. We included 428 patients (mean age: 49 ± 18 years, 262 males) with LVH (346 hypertrophic cardiomyopathy, 45 cardiac amyloidosis, 11 Anderson–Fabry disease, 16 endomyocardial fibrosis, 10 aortic stenosis). Our control group consisted of 234 healthy subjects (mean age: 35 ± 15 years; 126 males) without any known cardiovascular diseases. The developed machine-learning-based model achieved a 92% F1-score and 97% recall on the hold-out dataset, which is comparable to the medical experts. Experiments showed that the standardization method was able to significantly boost the performance of the algorithm. The algorithm could improve the diagnostic accuracy, and it could open a new door to AI applications in CMR.

Keywords: classification; left ventricular hypertrophy; CMR; machine learning



Citation: Budai, A.; Suhai, F.I.; Csorba, K.; Dohy, Z.; Szabo, L.; Merkely, B.; Vago, H. Automated Classification of Left Ventricular Hypertrophy on Cardiac MRI. *Appl. Sci.* **2022**, *12*, 4151. <https://doi.org/10.3390/app12094151>

Academic Editor: Julio Garcia Flores

Received: 16 March 2022

Accepted: 16 April 2022

Published: 20 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Cardiovascular diseases are the leading cause of death in developed countries [1,2]. Cardiovascular magnetic resonance (CMR) provides functional and morphological information of the heart for the evaluation, management, and diagnosis of patients with suspected or established cardiovascular disease. CMR is a multi-parametric, non-invasive imaging modality, which is considered the gold standard for the assessment of global and regional function and is able to evaluate myocardial perfusion and viability, tissue characterization, and coronary artery anatomy [3]. Left ventricular hypertrophy (LVH) is present in 15% to 20% of the population. It is more common in Afro-Americans and in patients with hypertension and obesity [4]. LVH is an independent predictor of future cardiovascular events, including coronary heart disease, heart failure, and stroke, regardless of its etiology [5,6]. The definition of LVH is an increase in left ventricular mass either due to an increase in wall thickness, an increase in cavity size, or both. In clinical practice, LVH is a common condition, which can be caused by diverse physiological and pathological mechanisms such as athlete's heart, hypertension, aortic stenosis, hypertrophic cardiomyopathy, infiltrative heart muscle disease, storage, and metabolic disorders (amyloidosis, Anderson–Fabry disease, etc.). LVH can develop silently over several years without symptoms, and it can be difficult to diagnose. The electrocardiogram (ECG) is a useful, but less sensitive tool for detecting LVH. The utility of the ECG lies in its relative inexpensiveness and wide availability. Its limitations stem from its moderate sensitivity or specificity, depending

on which of the various diagnostic criteria are applied [7,8]. In the Multi-Ethnic Study of Atherosclerosis patients who underwent MRI and ECG, it was found that various ECG criteria had low sensitivity for the detection of LVH [9]. As a result of these limitations of ECG, LVH is most reliably identified on imaging with echocardiography or CMR. Prior studies primarily used ECG [10] or M-Mode and two-dimensional (2D) echocardiography to identify LVH.

Conventional 2D echocardiography is the first-line imaging modality, which is used to evaluate the patterns, extent, and distribution of LVH and other anatomic and functional parameters and ventricular function. Nonetheless, echocardiography is limited by intra-observer and inter-observer variability, acoustic windows, and the lack of tissue characterization. Echocardiography-based LVH evaluation varies among the different definitions by ultrasound technicians and laboratories around the world, leading to inconsistency among epidemiological studies, and therefore, this could limit its clinical application [11]. CMR provides a comprehensive evaluation of myocardial hypertrophy regarding the extent and distribution of LVH and tissue characterization. Accurate measurements of wall thickness, the phenotype of hypertrophy, chamber size, and ventricular function can be obtained without any limiting factors such as imaging windows and body habitus. Importantly, CMR has a myocardial tissue characterization property that allows phenotypic determination of the LVH and careful evaluation of the precise etiology of LVH, which is a challenging clinical problem [12]. Hypertrophic cardiomyopathy (HCM) is the most common inherited cardiac disease, which often leads to sudden death in young people with an estimated prevalence of about 1:500 [13]. LVH is characterized by sudden cardiac death, stroke, and heart failure, but also decreased life expectancy [14,15]. The different LVH morphologic pattern can be precisely assessed by CMR and able to identify segmental hypertrophy, which can be difficult for echocardiography (i.e., apical HCM).

The application of machine learning methods on CMR images has boomed in the past 5–6 years. A huge body of work is available on automatic processing of MRI images, such as the segmentation of the ventricles [16–18], left ventricular quantification [19], pediatric cardiomyopathy classification [20], left ventricle wall motion classification [21], and cardiovascular event prediction for dilated cardiomyopathy [22]. These research endeavors also resulted in new architectures, developed for this field specifically, e.g., U-net [23] for myocardium segmentation, ν -net for cardiac vascular segmentation [24], and Ω -net for multiview CMR detection, orientation, and segmentation [25]. Besides the algorithmic improvement, the increasing availability of public datasets fuels the breakthroughs in the healthcare domain [26,27]. For LVH, the lack of publicly available benchmarks is reflected in the lower number of papers on the subject [28]. However, for echocardiography, there are promising works out there [29–31]. For instance, in paper [32], the echocardiography-based hypertrophy detection calculates the wall thickness, then the decision is made. This method automates the wall thickness measurements with thresholding, then the wall thickness is calculated from the adjusted contours. The accuracy of echocardiography-based diagnosis tends to be lower than that of CMR-based examinations [33,34]. This further motivates the application of CMR for hypertrophy detection. In the work of [35], the disease classification was based on a multi-stage process. First, it segments the heart, then it calculates the volumetric data. The volumetric data are used as features in a random forest to obtain the classification. The classes are different from ours, but there are HCM and normal heart cases, while the input images are CMR images. In paper [36], the authors developed an automatic wall thickness measurement on CMR images. The measurement was based on endocardial and epicardial segmentations. Another method relies on the clinical assessment of normal ranges for different morphological characteristics [37]. To the best of our knowledge, this is the first paper discussing an automatic method for left ventricular hypertrophy classification from CMR images.

If the algorithm can detect suspicious features of hypertrophy during a regular CMR examination, this indicates that the applied CMR protocol should be changed by adding necessary measurements and sequences on-site for a more detailed evaluation without the

need for additional examinations. During the post-process evaluation, it could improve the diagnostic accuracy by recognizing a milder, incipient form of LVH, which can be challenging for the less-experienced readers. The early detection of LVH and appropriate therapy will decrease cardiovascular morbidity and mortality [38]. In this paper, steps toward this ambition were made by developing an algorithm that considers more views of the heart and classifies the patient's hearts as normal or exhibiting hypertrophy. The algorithm we developed achieved results comparable to the human readers. Its high recall and sufficient precision allow for its use in an on-site setting, potentially causing the operators to change the CMR protocol (e.g., to administer the contrast agent, acquire late enhancement images, etc.) if hypertrophy is suspected. During the CMR examination, usually, the long-axis cine images are acquired first, then the short-axis cine images, then the late enhancement images if needed. We found that if the algorithm is restricted to only use long-axis cine images, it is still sufficient to alert the operator in order to select an appropriate CMR protocol, but might be limited in some selected cases. The rest of the paper is structured the following way: In Section 2, we introduce the dataset we utilized during our research, then we describe how our method works. In Section 3, we report the experimental results on a hold-out dataset and we make a comparison to the human-level performance. Section 4 describes our concluding thoughts.

2. Materials and Methods

The goal of this research is to develop an algorithm for hypertrophy classification from CMR scans. The scans contain more views: short-axis, long-axis. Our dataset was collected from the database of the The Heart and Vascular Center of Semmelweis University. Our method is based on the raw image scans with all available views, and the classification result is the direct output; we did not calculate intermediate features such as wall thickness.

2.1. Dataset

After the exclusion of patients with poor image quality, we investigated 428 patients (mean age: 49 ± 18 years, 262 males) with left ventricular hypertrophy in whom CMR examination was clinically indicated and 234 healthy subjects (mean age: 35 ± 15 years; 126 males) without any known cardiovascular diseases as a control group. The patients underwent CMR examination in our tertiary referral center between January 2009 and February 2019. Out of the 428 LVH patients, 346 had HCM (age: 46.9 ± 18.2 144 males), 45 patients had cardiac amyloidosis (age: 63.9 ± 9.7 years, 26 males), 11 patients had Anderson–Fabry disease (age: 48.3 ± 12.9 years, 7 males), 16 patients had endomyocardial fibrosis (age: 46.4 ± 14.3 , years 9 males), and 10 patients had aortic stenosis (age: 63.4 ± 17.5 years, 5 males). Appendix C shows example images. CMR examinations were performed on a 1.5 T magnetic resonance (MR) scanner (Achieva, Philips Medical Systems) using a cardiac coil. ECG gated balanced steady-state free precession (bSSFP) cine images were acquired in the three standard long-axis views: 2-chamber, 4-chamber, and LV outflow tract views. The protocol used for cine images in the present study was described in detail in a previous publication [39]. Short-axis (SA) images were also acquired with the full coverage of the left ventricle.

2.2. Model Architecture

The algorithm decides whether the patient has hypertrophy. The input to the algorithm is created from CMR scans of 4 views (axis). We used multi-view data because hypertrophy classification is challenging and different views provide different information. It is possible to see a pattern on a short-axis image that cannot be seen on the long-axis images or the other way around. The input images were collected from 4 views:

- Short-axis images from the apex to the base at different stages of the cardiac cycle;
- Long-axis, two-chamber images at different stages of the cardiac cycle (heart beat);
- Long-axis, three-chamber images at different stages of the cardiac cycle;
- Long-axis, four-chamber images at different stages of the cardiac cycle.

The usage of all the images from the short-axis scan has difficulties. The input would be too big, and the number of images were not the same for all patients. Therefore, for the short-axis view, we took three images in each second phase of the cardiac cycle. At each chosen cardiac cycle, we used one image from the basal, one from the mid, and one from the apical region, resulting in 36 images; see Figures 1 and 2. In the case of the long-axis views, we took each second image from the cardiac cycle, resulting in 12 images; see Figure 3.

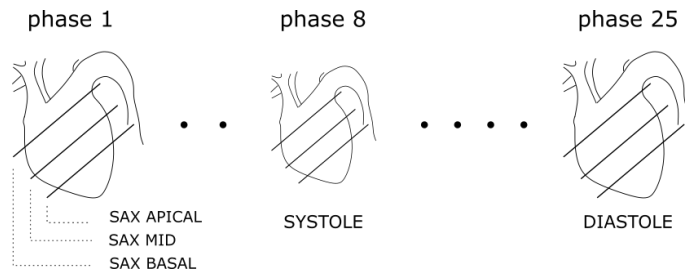


Figure 1. The illustration of how the short-axis images are created. The systole is the phase where the heart volume is the lowest, while in the diastole, it is the highest. A heart beat (cardiac cycle) is divided into 25 phases. In each phase, several slices are scanned. See the parallel lines. There are three major regions for the heart: basal, mid, apical.

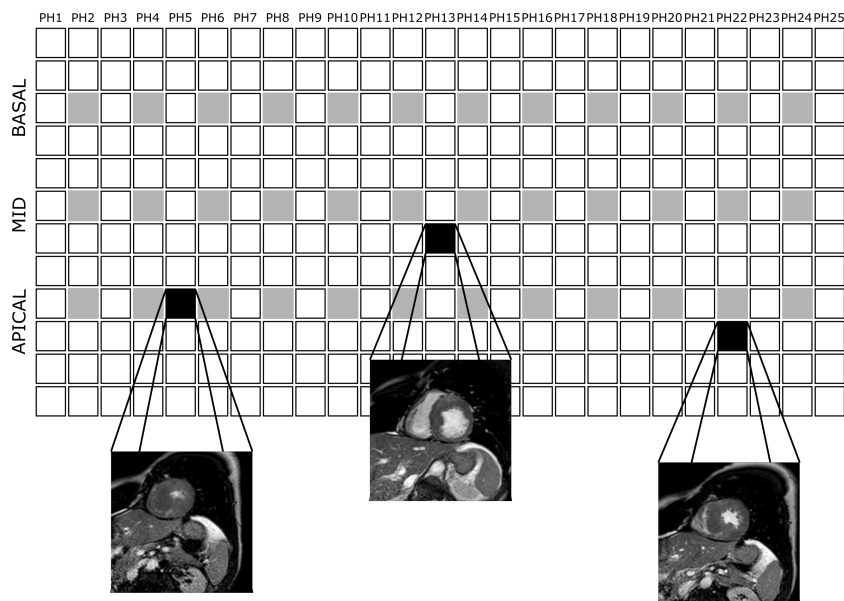


Figure 2. A short-axis CMR scan produces around 400 images for a heart. There are 12–16 slices visited by the scanner at each phase (time point). During one heart beat (cardiac cycle), 25 images are created for a slice. There are three main regions: apical, mid, basal. See also Figure 1. To create a fixed-size input for the model, we picked three images from each second phase; see the gray boxes. Overall, this results in 36 images. Black squares show example images, how the real image looks at a given slice and phase.

The model is an ensemble of the extractors of the separate views. Images from each view are fed into a separate network to extract features. The features are concatenated, then the ensemble classifier is applied to obtain the prediction (normal or exhibiting hypertrophy); see Figure 4. The architecture of the extractor for the best-performing model can be seen in Figure 4, left side. We used the same extractor for each view. The extractors were trained separately; therefore, a temporary layer was applied to create a temporary classifier. The architecture of the temporary layer can be seen in Table 1. After the extractors are trained, an ensemble model is created with an ensemble classifier; see Figure 4, right side, under the classifier block. The models were built from residual blocks, with each block containing 3-dimensional convolutions and batch normalizations; see Figure 4, bottom part.

The reason for the 3D convolution is the positive effect of considering the time dimension of the input (how the heart moves). For further elaboration on the performance and the choices we made, see the details in Section 3.3.



Figure 3. Long-axis CMR scan has three view: 2-chamber, 4-chamber, and 3-chamber views. For each view, 25 images are produced. The images are created from the same slice of the heart, but at different time points of the cardiac cycle. To create a fixed-size input, we picked an image from every second phase; see the gray boxes. Overall, this results in 12 images for each view (LA2, LA4, LA3).

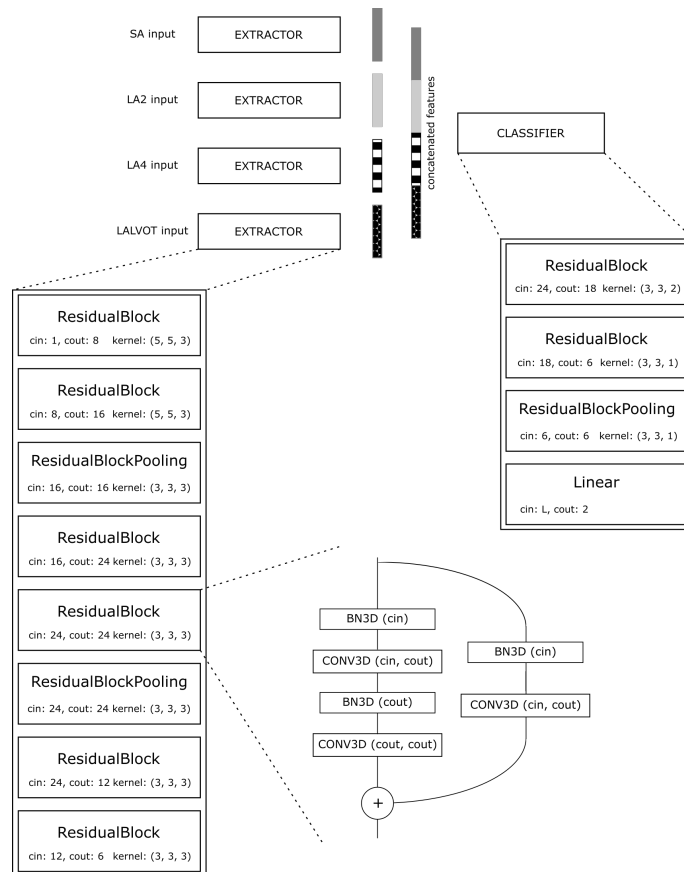


Figure 4. The schematic architecture of the model. Each view has an extractor and the output features from the extractors aggregated by concatenating the features along the channel dimension. This is the ensemble of the views. The extractor and the classifier are built from 3D residual blocks. The architecture of the 3D ResNet blocks can be seen in the middle of the image. The ResidualBlock and the ResBlockPooling differ in the strides. For the pooling block, the first convolution and the convolution on the skip branch have stride 2; otherwise, it is 1. The activations are ReLUs, which were applied after the batch normalization layers. In the ResidualBlock version, we applied padding in each convolution, while in pooling, we applied padding in the last convolution of the straight branch. Padding was: $(k - 1)/2$ for each dimension. The kernel sizes were chosen as odd values in each case.

Table 1. The architecture of the temporary classifier. After the pooling layer, the tensors are reshaped to (batch, L) size. The value of L is different for the long-axis and short-axis views.

Block	Block Name	Cin	Cout	Kernel Size
TC1	ResBlockPooling	6	3	(3, 3, 1)
TC2	Linear	L	2	-

2.3. Preprocessing and Data Augmentation

Before the images are fed into the model, two main steps are executed: (1) preprocessing and (2) augmentation. The augmentation is the same for each of the views, but preprocessing contains an additional step for the long-axis views. Preprocessing always applies noise reduction by cropping the intensity values between the 1st and 99th percentiles. Then, the images are normalized into a 0–1 interval. For the long-axis views, the images are standardized because their orientation shows high variance. Standardization is achieved by a superposition to a reference frame calculated for each view separately. The reference frame is given as the normal vector of a typical image for a given view. The superposition applies mirroring and a rotation around the center point of the image to be preprocessed. Appendix A gives further insight into the details of the standardization. In Section 3.3, further details are shown about the effect of the standardization on performance. The augmentation contains a random rotation and Gaussian noise.

2.4. Training Scheme

We trained the model in two stages. This training process falls into the supervised learning paradigm, because we have the ground truth pathologies for each scan. The dataset was unbalanced; therefore, we sampled the normal group with higher probability to equalize the occurrences of hypertrophic and normal samples in the training batches; see Section 2.1 for the ratio. The dataset was split into three parts: training (70%), validation (15%), and testing (15%). The test set was created only once, and we kept it until the final test with the best model chosen on the validation test. We repeated the training with each parameter setting three times to understand the stability of the results. In each repetition, the training and validation parts were resampled. First, the feature extractors were trained separately to predict whether the patient had hypertrophy. For this part, we used a temporary layer at the end of each extractor to create a classifier. Then, the temporary layer was removed, and the ensemble model was built. For combining the outputs of the feature extractors, we concatenated (the long-axis was padded in the depth dimension) the features, then fed them into the classifier; see Figure 4. The whole ensemble was trained, but the feature extractors' weights were frozen. The training was applied on different combinations of the possible views. The combinations were based on realistic scenarios, because the earlier we can detect the condition of hypertrophy, the faster the operators can react during the scanning procedure. In a clinical setting, the examination process mostly follows similar orders among the views. During the CMR examination, the typical order was long-axis views, then short-axis view. It is important to test only using the long-axis view, the short-axis view, and then, their combination. The parameters of the best model can be seen in Table A1.

2.5. Human Evaluation

The performance of the algorithm was also compared to human experts (hereafter readers). The design of the evaluation simulated a realistic setup for an everyday examination procedure. The readers were asked to read CMR scans of 117 subjects, but they were not told the real purpose of the study. About each subject, a very brief patient history was provided (without giving clear reference to the real disease) along with the images of a full MRI scan. This included the short-axis and long-axis images. For the analysis, we included CMR scans from the normal group as well and the following pathologies: acute or chronic myocardial infarction, dilated cardiomyopathy, Takotsubo cardiomyopathy, and acute

myocarditis. The list contained the most frequent pathologies encountered during regular assessments. We also included different cardiac pathologies that could cause LVH (HCM, Anderson–Fabry disease, amyloidosis, aortic stenosis, and endomyocardial fibrosis). The reason for pathologies outside of hypertrophy was to avoid bias during the evaluation. Overall, six experts finished the experiment. Two of them were senior colleagues (25 and 10 years of experience) and three of them at the mid-senior level (4–7 years of experience), and one of them was a junior (2 years of experience).

3. Results

We experimentally proved that the algorithm described in Section 2 can achieve comparable performance to human experts.

3.1. Results of Human-Evaluation

The human evaluation established a baseline to raise expectation against the algorithm. Table 2 shows the results. Overall means the accuracy of the diagnosis of each expert for all 117 subjects. This includes all the pathologies. In the Hyp-Norm row, the pathologies are grouped into two groups, normal and hypertrophy, which includes all the LVH etiologies considered earlier in this paper. The prediction of a reader was considered as valid if the predicted pathology fell into the hypertrophy group, but the etiology did not have to be accurate. In the HCM row, we measured the accuracy of differentiating between the patients with HCM and other cardiac disorders, which usually represents LVH. In the last three rows, precision, recall, and F1-score were calculated for the Hyp-Norm case. Hypertrophy was considered as a positive event in the confusion matrix. If we compare the consistency among the experts in terms of three groups: normal, hypertrophy, and the rest, we found 83 %, 71 %, and 91 % consistency values, respectively. Consistency is defined as an agreement among at least five radiologists. The high value of recall and the lower value of consistency for the normal group indicates that radiologists tend to classify healthy patients as those having a condition. This is understandable, as a false positive can easily prove to be negative after some further examinations. On the contrary, false negatives can lead to delayed and inappropriate patient care.

Table 2. The scores of the human evaluation. The scores are consistent across the readers; the variance is small.

Scores	R1	R2	R3	R4	R5	R6	Mean
Overall	89.7	91.5	91.5	88.0	89.7	90.6	90.0
Hyp-Norm	92.8	95.7	91.3	94.2	89.9	94.2	93.0
HCM	85.4	89.6	91.7	85.4	93.8	85.4	88.6
Precision	97.8	97.9	88.9	92.3	87.3	95.8	93.3
Recall	93.6	95.8	100	100	100	97.9	97.9
F1	95.7	96.8	94.1	96.0	93.2	96.8	95.4

3.2. Performance of the Algorithm

The performance of the best model can be seen in Table 3. The table shows that only using the LA views was enough to achieve comparable results to humans by considering the standard deviations as well (3–4%). This is important, because the contrast agent can be injected after the long-axis measurements (if the algorithm indicates it and the experts accept it), then the short-axis cine images can be acquired, since the late enhancement images could be acquired at least 10 min after contrast material administration. This approach can save significant amounts of time and can also warn the on-site medical staff that the MRI protocol should be changed in order to avoid further, unnecessary examinations.

Table 3. The performance of the best model on the test sets. The last row shows the result on the validation set when the long-axis views were combined with the short-axis views.

Cases	F1	Precision	Recall
only LA2	90	86	92
only LA4	86	81	90
only LA3	91	86	92
only SA	86	90	83
all LAs	89	84	90
LA+SA	91	88	96
v. LA+SA	93	91	94

The box plots in Figures 5 and 6 were calculated by repeating the test evaluation on 20 randomly sampled subsets of the test data, and in each sample, we used 70% of the test data. This method is similar to bootstrapping. Both images show the same relative performance. The algorithm using only the LA views had lower performance, but when short-axis and long-axis views were combined, the human level and the algorithm scores became close to each other, especially in the case of the recall. The results showed lower F1 and recall for the only short-axis case (see Table 3), which can be a result of the higher complexity of the data. More samples for the SA case could scale up the performance. Similarly, the algorithm (SA+LA) had lower performance than the experts, but we claim that a larger dataset would reduce the gap.

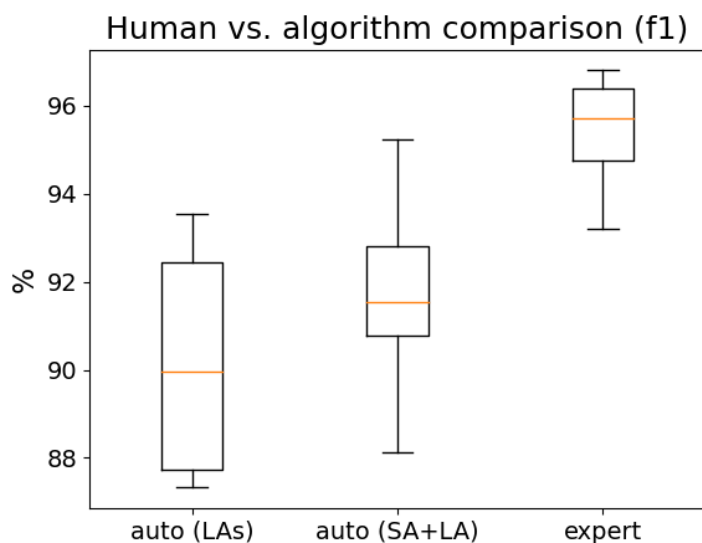


Figure 5. Comparison of the human (expert) and algorithm (auto) performances. The *p*-value between auto (LA) and auto (SA+LA) is lower than 0.001, which means using the short-axis images contributed to a significantly better performance. Between the auto (SA+LA) and expert group, the *p*-value was less than 0.001. For calculating the *p*-values, we used two-sample *t*-tests.

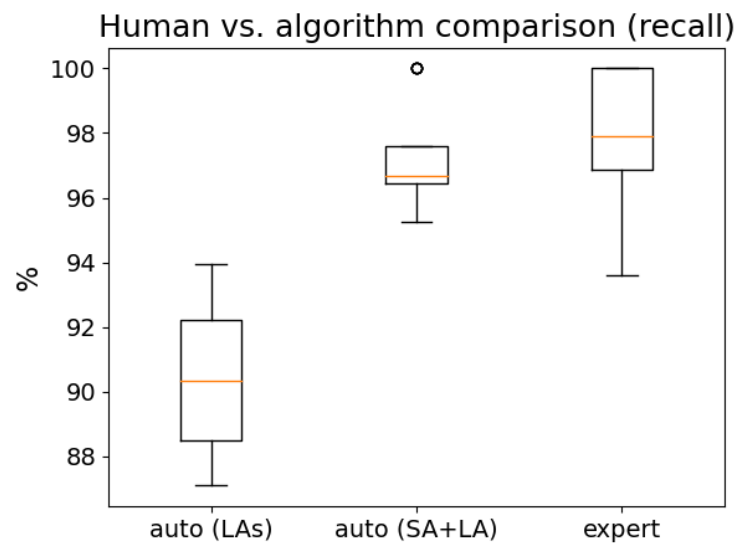


Figure 6. Comparison of the human (expert) and algorithm (auto) performances. High recall is beneficial because the algorithm can identify samples suspicious of hypertrophy with a high probability. The false positives can be handled by the experts who supervise the examination. The p -value is less than 0.001 between the auto (LA) and auto (SA+LA) groups. When comparing auto (SA+LA) and the expert groups, we obtained a p -value = 0.3, indicating there was no statistically significant difference between them. Therefore, the auto (SA+LA) was statistically identical to the expert group in terms of the recall.

3.3. Ablation Study

We executed several experiments before we arrived at the final model, data processing, and parameter choices. In this subsection, we briefly summarize our findings. We cover the three main aspects of the algorithm:

1. Model selection;
2. Data preprocessing;
3. Hyper-parameter setting.

The above order does not represent the order of our experiments. It was established in order to explain our experience in more logical fashion. We did not measure every possible combination of choices; therefore, we can explain and showcase the tendencies of the different choices.

Model selection. We tried three main architectures. The first architecture was a fully convolutional model with 4–5 convolutional layers, assuming the ensemble model with more views can achieve good results overall and we would not need strong learners per view. Our results indicated that bigger networks would be required to achieve scores (accuracy, F1-score, etc.) around 90 percent. The second architecture was similar to ResNet with two-dimensional convolutions. The time dimension in the long-axis view was stacked together to form a 12-channel image. The structure was similar to the ResNet described in Section 2.2. We experienced significant performance growth (around 3–4 percent) as the model size achieved 8 residual blocks, meaning 16 convolutional layers overall. Further increasing the size did not affect performance significantly. One reason for that may be the size of the dataset. During the data-preprocessing-related changes, we came to the conclusion that taking into account the time dimension (basically the movement or dynamic patterns of the heart) had a major effect on the results (over six percent in the case of the short-axis views). Therefore, we created a 3D convolution-based ResNet model to properly handle the time dimension. We formed a 3D image, as time became the depth dimension of the image. This model performed better and more robustly (regarding the sensitivity for the hyper-parameters). However, the drawback of the 3D ResNet lies in its slow training

speed. As the performance on the short-axis view was worse, we tried to increase the model size for this view only, but this did not cause relevant changes. Finally, we used the same architecture for all the views.

Data preprocessing. Data preprocessing and the input representation to the network proved to be the most important factors. To speed up the training, we tried less input data first. We used only two images from the long-axis views, one from the systole phase and one from the diastole phases. We used six images from the short-axis view and three images at the systole and the diastole phases, respectively. This input formation resulted in fair accuracy values (around 84 percent), but it turned out that taking images from other points of the cardiac cycle contributed to better results. Standardization (see Appendix A) had a very important role in achieving the final results. We identified the long-axis views to be noisy as a result of the different orientations of the images. This was not true for the short-axis. One way to cope with this is to use random rotation for augmentation with degrees between 0 and 180. We found this approach to be inefficient in helping the learning process. The standardization method caused a significant performance growth. Therefore, we used only a small eight-degree angle for rotation during augmentation. We also used cropping and some noise during augmentation.

Hyper-parameter tuning. When a model and a data preprocessing method were chosen, there were some hyper-parameters to optimize. These were batch size, number of epochs, learning rate, optimization algorithm, loss function, regularization method and their parameters, and the cropping size of the image. We chose batch size 16 because 8 was too noisy for the training. Larger batch sizes require too much memory. The number of epochs was chosen between 20 and 50, and we used early stopping to avoid overfitting. We found that the AdamW [40] algorithm with learning rate 5×10^{-4} achieved better results than Adam, SGD, and RMSProp. We used focal loss [41], because focal loss can distinguish the easy samples from the difficult ones by applying a factor $((1 - p)^\gamma)$, which reduces the loss for the well-classified samples. Our intuition was that the samples contained some very difficult cases (due to etiologies such as amyloidosis, which is difficult to diagnose), and therefore, focal loss could help. In our experiments, we experienced L1 and L2 losses to be harmful, and dropout with large values was disadvantageous. This can be explained by the observation that batch normalization has some regularization effect, which can eliminate the need for dropout [42], and our 3D ResNet contains batch normalization layers. The final cropping size of the input image proved to be 150×150 . Smaller (120×120) and larger sizes (190×190) were worse. For the larger size, the image can contain too much noise, while the smaller crop can miss some details with the heart not always being at the center of the image.

4. Discussion and Conclusions

Cardiovascular diseases are the leading causes of death around the world [1,2,43]. LVH is a well-recognized independent risk factor for several cardiovascular complications [5]. The diagnosis of LVH can be challenging. For this, there are some methods used in clinical practice such as electrocardiography, echocardiography, and CMR. CMR is a non-invasive tool for diagnosing myocardial pathologies. CMR-based hypertrophy detection can be more efficient and reliable and may improve the diagnostic method in order to recognize LVH in an earlier stage. We developed a deep-learning-based algorithm for identifying left ventricular hypertrophy during a CMR examination (on-site) and for helping the diagnostic process following the examination (off-site). The on-site application can save time, if the algorithm indicates the presence of LVH right after the long-axis measurements; therefore, some additional, necessary images could be acquired and contrast administration should be applied. With the use of on-site application, the CMR protocol can be changed during the scanning, in order to avoid the need to call back the patient for an additional CMR examination to provide the correct diagnosis. Nevertheless, if the algorithm is used during post-process evaluation, it can warn the reader that LVH is present, so the diagnostic accuracy can be improved. This is important because the identification of the incipient

or milder form of LVH is difficult for less-experienced readers, and early detection of LVH and subsequent therapy are key factors in reducing cardiovascular morbidity and mortality [38,44]. Our algorithm achieved a performance close to the medical experts' (readers) scores. Our comparison was based on the F1-score, precision, and recall. The model we implemented was an ensemble model. Each view had a separate extractor, and the features extracted from the acquired images were concatenated. Then, an ensemble classifier takes the concatenated features as the input and calculates the probability of having LVH. The dataset was collected from the Heart and Vascular Center of Semmelweis University, and it contains the raw image scans with all available views (long-axis and short-axis cine images) and the corresponding pathologies.

Our algorithm had a recall rate of 90% when the combination of long-axis views was used as the input. In the case of the combination of long-axis views and short-axis views, we had a 96% rate. The corresponding F1-scores were 89% and 91%, respectively. High recall is beneficial, because fewer LVH cases will be left undiagnosed. False positives (predicted as LVH, yet normal) can be discarded by the experts supervising the examination. In order to judge the applicability of our method, we established a baseline by measuring the scores of medical experts. The measurement involved six readers with varying levels of experience. The measurement was designed to simulate a realistic clinical scenario where the reader has no clear reference to the real case, but has access to the images of full CMR scans. To make it more realistic, we included several other diseases in addition to LVH. We included diseases that appear frequently in clinical practice, and the readers were blinded to the purpose of the study. There are three main outcomes of the human experiment: (1) the differences among the scores (F1-score, recall, etc.) of the readers were surprisingly small; (2) recall was the highest value indicating that the readers had a bias toward having a cardiac disease; (3) we obtained the baseline values for the scores (F1-score—95%, recall—98%); see Table 2. High recall was also achieved by our algorithm in the case of the combined long-axis and short-axis model. Figures 5 and 6 indicate that our algorithm can already be advantageous in clinical practice even though there is still room for improvement.

We claim that by using a larger dataset, the gap can be bridged and that this method can be a good candidate to become part of the daily clinical routine during CMR examinations. Our method was limited to only one vendor and clinic center. For creating a more robust method, the model should be trained on data gathered from different clinic centers and vendors. Another limitation is the classification of etiologies. The current method differentiates between two groups, normal (healthy) subjects and hypertrophy. There are different etiologies for hypertrophy (e.g., HCM, amyloidosis), which can be differentiated by including late enhancement images. From the dataset, we excluded healthy athletes, but LVH can be present as a physiological condition in athlete's heart; therefore, it could be an interesting topic to differentiate between physiologic and pathologic LVH.

To the best of our knowledge, this is the first paper where a method for automatic classification of LVH from different CMR images (short-axis, long-axis cine images) was investigated and compared to medical experts. Future work can focus on the separation of the etiologies within LVH automatically. Sports-related LVH should be also addressed in order to create a more complete methodology.

Author Contributions: Conceptualization, A.B., F.I.S. and H.V.; methodology, A.B. and F.I.S.; software, A.B.; validation, A.B. and F.I.S.; resources, K.C., H.V. and B.M.; data curation, A.B., F.I.S., Z.D. and L.S.; writing—original draft preparation, A.B. and F.I.S.; writing—review and editing, A.B., F.I.S., K.C. and H.V.; visualization, A.B. and F.I.S.; supervision, K.C., H.V. and B.M.; funding acquisition, K.C., H.V. and M.B. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Research, Development and Innovation Office of Hungary (NKFIA; 019-1.1.1-PIACI-KFI-2019-00263). This study was also supported by the National Research, Development and Innovation Office of Hungary (NKFIA; NVKP-16-1-2016-0017). The study was financed by the Research Excellence Programme of the Ministry for Innovation and Technology in Hungary within the framework of the Bioimaging Thematic Programme of Semmelweis University and by the Ministry of Innovation and Technology NRD Office within the framework of the Artificial

Intelligence National Laboratory Program. Project no. TKP2021-NKTA-46 has been implemented with the support provided by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021-NKTA funding scheme.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of The Heart and Vascular Center of Semmelweis University (Protocol Code NVKP, date: 5 June 2020).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author. The data are not publicly available; for acquiring the dataset, the permission of the local institution board is necessary.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Standardization

The method for standardizing the long-axis images is based on defining a reference system relative to a fixed axis. This axis is the Z-axis, which points from the feet to the head of the patient, running parallel with the bore of the MRI. Each acquired image has a plane, and it can be described in this coordinate system. The plane is characterized by its normal vector and its orientation. The orientation is relative to the Z-axis. The images are stored in dicom files, which contain the orientation and position matrices. The standardization is achieved by rotation with a proper angle around the axis parallel with the normal vector and crossing the middle of the image; see Figure A1.

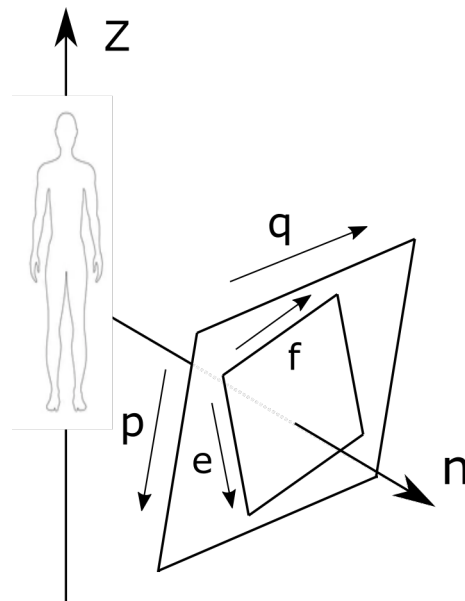


Figure A1. Schematic illustration of the vectors used in the explanation of the standardization process.

First, the algorithm calculates the normal vector of the image from the orientation matrix. The orientation matrix contains the directions of the left side and the upper side of the image (\vec{e} and \vec{f}). Therefore, the normal vector is:

$$\vec{n} = \vec{e} \times \vec{f}. \quad (\text{A1})$$

The normal vectors are almost the same for each view. Then, a new reference frame can be calculated (\vec{p} and \vec{q}):

$$\vec{q} = \vec{z} \times \vec{n}, \quad \vec{p} = \vec{q} \times \vec{n} \quad (\text{A2})$$

where $\vec{z} = (0, 0, 1)$, then \vec{p}, \vec{q} are normalized. The orientation is defined as the direction of \vec{e} in the \vec{p}, \vec{q} plane:

$$\vec{d} = [\vec{e} \cdot \vec{p}, \vec{e} \cdot \vec{q}]. \quad (\text{A3})$$

We can define a reference orientation (\vec{d}_0), then each image can be compared and rotated against the reference orientation. To decrease the size of the required rotation angle, we calculated the average orientation of the images in the dataset per view. Then, we defined the reference orientations according to the average values. For the sake of completeness, these values were: LA2 ($-0.937, 0.166$), LA4 ($0.632, 0.032$) and LALVOT ($-0.0054, -0.635$). The rotation angle (φ) is given as follows:

$$\cos \varphi = \vec{d} \cdot \vec{d}_0. \quad (\text{A4})$$

Appendix B. Parameters

Table A1. The hyper-parameters used in the best-performing model.

Parameter	Value
batch size	16
learning rate	0.0005
optimizer	AdamW
input shape	150×150
max angle	8

Appendix C. Example Images

The following images show examples for different heart conditions: normal, HCM, amyloidosis, and Anderson–Fabry disease. In each row of pictures, the views from left to right are the following: short-axis, long-axis 2-chamber, long-axis 4-chamber, and long-axis 3 chambers-view.

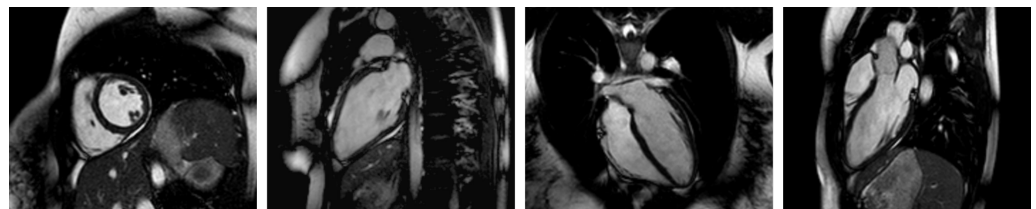


Figure A2. Short-axis cine image and long-axis cine images of healthy subject without left ventricular hypertrophy.

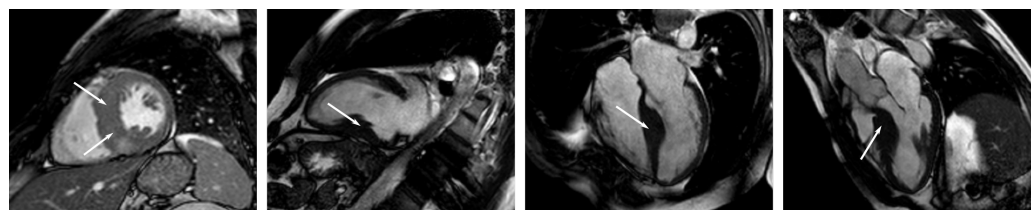


Figure A3. Short-axis cine image and long-axis cine images demonstrate left ventricular hypertrophy in a patient with hypertrophic cardiomyopathy (HCM). Cine images show marked asymmetrical septal hypertrophy (white arrows) corresponding with HCM.

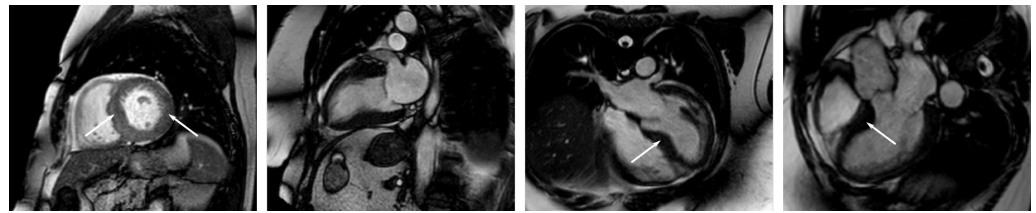


Figure A4. Short-axis cine image and long-axis cine images demonstrate concentric left ventricular hypertrophy with subtle septal predominance (white arrows), in a patient with Anderson–Fabry disease.

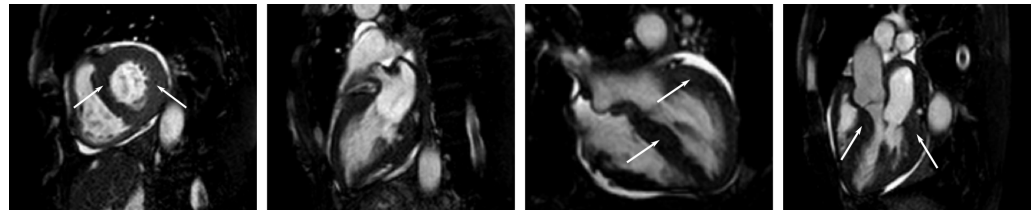


Figure A5. Short-axis cine image and long-axis cine images show marked, concentric left ventricular hypertrophy (white arrows) in a patient with cardiac amyloidosis.

References

- Namara, K.; Alzubaidi, H.; Jackson, J.K. Cardiovascular disease as a leading cause of death: How are pharmacists getting involved? *Integr. Pharm. Res. Pr.* **2019**, *8*, 1–11. [[CrossRef](#)] [[PubMed](#)]
- Roth, G.A.; Mensah, G.A.; Johnson, C.O.; Addolorato, G.; Ammirati, E.; Baddour, L.M.; Barengo, N.C.; Beaton, A.Z.; Benjamin, E.J.; Benziger, C.P.; et al. Global Burden of Cardiovascular Diseases and Risk Factors. *J. Am. Coll. Cardiol.* **2021**, *77*, 1958–1959. [[CrossRef](#)]
- Foley, J.R.; Plein, S.; Greenwood, J.P. Assessment of stable coronary artery disease by cardiovascular magnetic resonance imaging: Current and emerging techniques. *World J. Cardiol.* **2017**, *9*, 92–108. [[CrossRef](#)] [[PubMed](#)]
- Cuspidi, C.; Sala, C.; Negri, F.; Mancia, G.; Morganti, A. Prevalence of left ventricular hypertrophy in hypertension: An updated review of echocardiographic studies. *J. Hum. Hypertens.* **2012**, *26*, 343–349. [[CrossRef](#)]
- Bluemke, D.A.; Kronmal, R.A.; Lima, J.A.; Liu, K.; Olson, J.; Burke, G.L.; Folsom, A.R. The relationship of left ventricular mass and geometry to incident cardiovascular events: The MESA (Multi-Ethnic Study of Atherosclerosis) study. *J. Am. Coll. Cardiol.* **2008**, *52*, 2148–2155. [[CrossRef](#)]
- Kawel-Boehm, N.; Kronmal, R.; Eng, J.; Folsom, A.; Burke, G.; Carr, J.J.; Shea, S.; Lima, J.A.; Bluemke, D.A. Left Ventricular Mass at MRI and Long-term Risk of Cardiovascular Events: The Multi-Ethnic Study of Atherosclerosis (MESA). *Radiology* **2019**, *293*, 107–114. [[CrossRef](#)]
- Devereux, R.B. Is the electrocardiogram still useful for detection of left ventricular hypertrophy? *Circulation* **1990**, *81*, 1144–6. [[CrossRef](#)]
- Goldberger, A.; Goldberger, Z.; Shvilkin, A. *Goldberger's Clinical Electrocardiography: A Simplified Approach*, 9th ed.; Elsevier: Amsterdam, The Netherlands, 2017.
- Jain, A.; Tandri, H.; Dalal, D.; Chahal, H.; Soliman, E.; Prineas, R.; Folsom, A.; Lima, J.; Bluemke, D. Diagnostic and prognostic utility of electrocardiography for left ventricular hypertrophy defined by magnetic resonance imaging in relationship to ethnicity: The Multi-Ethnic Study of Atherosclerosis (MESA). *Am. Heart J.* **2010**, *159*, 652–658. [[CrossRef](#)]
- Okin, P.; Devereux, R.; Harris, K.; Jern, S.; Kjeldsen, S.; Julius, S.; Edelman, J.; Dahlöf, B.; Investigators, L.S. Regression of electrocardiographic left ventricular hypertrophy is associated with less hospitalization for heart failure in hypertensive patients. *Ann. Intern. Med.* **2007**, *147*, 311–319. [[CrossRef](#)]
- Foppa, M.; Duncan, B.; Rohde, L. Echocardiography-based left ventricular mass estimation. How should we define hypertrophy? *Cardiovasc. Ultrasound* **2005**, *3*, 17. [[CrossRef](#)]
- Grajewski, K.G.; Stojanovska, J.; Ibrahim, E.S.H.; Sayyoub, M.; Attili, A. Left Ventricular Hypertrophy: Evaluation With Cardiac MRI. *Curr. Probl. Diagn. Radiol.* **2020**, *49*, 460–475. [[CrossRef](#)] [[PubMed](#)]
- Semsarian, C.; Ingles, J.; Maron, M.; Maron, B. New perspectives on the prevalence of hypertrophic cardiomyopathy. *J. Am. Coll. Cardiol.* **2015**, *65*, 1249–1254. [[CrossRef](#)] [[PubMed](#)]
- Maron, B.J.; Maron, M.S. Hypertrophic cardiomyopathy. *Lancet* **2013**, *381*, 242–255. [[CrossRef](#)]
- Maron, B.J.; Ommen, S.R.; Semsarian, C.; Spirito, P.; Olivetto, I.; Maron, M.S. Hypertrophic Cardiomyopathy: Present and Future, With Translation Into Contemporary Cardiovascular Medicine. *J. Am. Coll. Cardiol.* **2014**, *64*, 83–99. [[CrossRef](#)]
- Bai, W.E.A. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *J. Cardiovasc. Magn. Reson.* **2018**, *20*, 65. [[CrossRef](#)]

17. Bernard, O.; Lal, E.A.; Zotti, C.; Cervenansky, F.; Yang, X.; Heng, P.A.; Cetin, I.; Lekadir, K.; Camara, O.; Ballester, M.A.G.; et al. Deep Learning Techniques for Automatic MRI Cardiac Multi-structures Segmentation and Diagnosis: Is the Problem Solved? *IEEE Trans. Med. Imaging* **2018**, *37*, 2514–2525. [[CrossRef](#)] [[PubMed](#)]
18. Lu, Y.; Connelly, K.; Dick, A.; Wright, G.; Radau, P. Automatic functional analysis of left ventricle in cardiac cine MRI. *Quant. Imaging Med. Surg.* **2013**, *3*, 200–209. [[CrossRef](#)]
19. Tao, Q.; Yan, W.; Wang, Y.; Paiman, E.H.; Shamoin, D.P.; Garg, P.; Plein, S.; Huang, L.; Xia, L.; Sramko, M.; et al. Deep Learning-based Method for Fully Automatic Quantification of Left Ventricle Function from Cine MR Images: A Multivendor, Multicenter Study. *Radiology* **2019**, *290*, 81–88. [[CrossRef](#)]
20. Gopalakrishnan, V.; Menon, P.G.; Madan, S. cMRI-BED: A novel informatics framework for cardiac MRI biomarker extraction and discovery applied to pediatric cardiomyopathy classification. *BioMed. Eng. OnLine* **2015**, *14*, S7. [[CrossRef](#)]
21. Mantilla, J.; Garreau, M.; Bellanger, J.J.; Paredes, J.L. Machine Learning Techniques for LV Wall Motion Classification Based on Spatio-temporal Profiles from Cardiac Cine MRI. In Proceedings of the 2013 12th International Conference on Machine Learning and Applications, Miami, FL, USA, 4–7 December 2013; Volume 1, pp. 167–172. [[CrossRef](#)]
22. Chen, R.; Lu, A.; Wang, J.; Ma, X.; Zhao, L.; Wu, W.; Du, Z.; Fei, H.; Lin, Q.; Yu, Z.; et al. Using machine learning to predict one-year cardiovascular events in patients with severe dilated cardiomyopathy. *Eur. J. Radiol.* **2019**, *117*, 178–183. [[CrossRef](#)]
23. Curiale, A.H.; Colavecchia, F.D.; Kaluza, P.; Isoardi, R.A.; Mato, G. Automatic myocardial segmentation by using a deep learning network in cardiac MRI. In Proceedings of the 2017 XLIII Latin American Computer Conference (CLEI), Córdoba, Argentina, 4–8 September 2017; pp. 1–6. [[CrossRef](#)]
24. Winther, H.B.; Hundt, C.; Schmidt, B.; Czerner, C.; Bauersachs, J.; Wacker, F.; Vogel-Claussen, J. v-net: Deep Learning for Generalized Biventricular Cardiac Mass and Function Parameters. *arXiv* **2017**, arXiv:1706.04397.
25. Vigneault, D.M.; Xie, W.; Ho, C.Y.; Bluemke, D.A.; Noble, J.A. Omega-Net: Fully Automatic, Multi-View Cardiac MR Detection, Orientation, and Segmentation with Deep Neural Networks. *arXiv* **2017**, arXiv:1711.01094.
26. Radau, P.; Lu, Y.; Connelly, K.; Paul, G.; Dick, A.; Wright, G. Evaluation Framework for Algorithms Segmenting Short Axis Cardiac MRI. *MIDAS J.* **2009**, *49*. [[CrossRef](#)]
27. Suinesiaputra, A.; Cowan, B.R.; Al-Agamy, A.O.; Elattar, M.A.; Ayache, N.; Fahmy, A.S.; Khalifa, A.M.; Medrano-Gracia, P.; Jolly, M.P.; Kadish, A.H.; et al. A collaborative resource to build consensus for automated left ventricular segmentation of cardiac MR images. *Med. Image Anal.* **2014**, *18*, 50–62. [[CrossRef](#)] [[PubMed](#)]
28. Martin-Isla, C.; Campello, V.M.; Izquierdo, C.; Raisi-Estabragh, Z.; Baeßler, B.; Petersen, S.E.; Lekadir, K. Image-Based Cardiac Diagnosis With Machine Learning: A Review. *Front. Cardiovasc. Med.* **2020**, *7*, 1. [[CrossRef](#)] [[PubMed](#)]
29. Madani, A.; Ong, J.R.; Tibrewal, A.; Mofrad, M.R.K. Deep echocardiography: Data-efficient supervised and semi-supervised deep learning towards automated diagnosis of cardiac disease. *Npj. Digit. Med.* **2018**, *1*, 59. [[CrossRef](#)] [[PubMed](#)]
30. Kwon, J.M.; Jeon, K.H.; Kim, H.M.; Kim, M.J.; Lim, S.M.; Kim, K.H.; Song, P.S.; Park, J.; Choi, R.K.; Oh, B.H. Comparing the performance of artificial intelligence and conventional diagnosis criteria for detecting left ventricular hypertrophy using electrocardiography. *EP Eur.* **2020**, *22*, 412–419. [[CrossRef](#)]
31. Jothiramalingam, R.; Jude, A.; Patan, R.; Ramachandran, M.; Duraisamy, J.H.; Gandomi, A.H. Machine learning-based left ventricular hypertrophy detection using multi-lead ECG signal. *Neural Comput. Appl.* **2021**, *33*, 4445–4455. [[CrossRef](#)]
32. Jian, Z.; Wang, X.; Zhang, J.; Wang, X.; Deng, Y. Diagnosis of left ventricular hypertrophy using convolutional neural network. *BMC Med. Inform. Decis. Mak.* **2020**, *20*, 243. [[CrossRef](#)]
33. De la Garza-Salazar, F.; Romero-Ibarguengoitia, M.E.; Rodriguez-Diaz, E.A.; Azpiri-Lopez, J.R.; González-Cantu, A. Improvement of electrocardiographic diagnostic accuracy of left ventricular hypertrophy using a Machine Learning approach. *PLoS ONE* **2020**, *15*, e0232657. [[CrossRef](#)]
34. Massera, D.; McClelland, R.L.; Ambale-Venkatesh, B.; Gomes, A.S.; Hundley, W.G.; Kawel-Boehm, N.; Yoneyama, K.; Owens, D.S.; Garcia, M.J.; Sherrid, M.V.; et al. Prevalence of Unexplained Left Ventricular Hypertrophy by Cardiac Magnetic Resonance Imaging in MESA. *J. Am. Heart Assoc.* **2019**, *8*, e012250. [[CrossRef](#)] [[PubMed](#)]
35. Khened, M.; Alex, V.; Krishnamurthi, G. Densely Connected Fully Convolutional Network for Short-Axis Cardiac Cine MR Image Segmentation and Heart Diagnosis Using Random Forest. Statistical Atlases and Computational Models of the Heart. In *ACDC and MMWHS Challenges*; Pop, M., Sermesant, M., Jodoin, P.M., Lalande, A., Zhuang, X., Yang, G., Young, A., Bernard, O., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 140–151.
36. Augusto, J.B.; Davies, R.H.; Bhuvu, A.N.; Knott, K.D.; Seraphim, A.; Alfarih, M.; Lau, C.; Hughes, R.K.; Lopes, L.R.; Shiwani, H.; et al. Diagnosis and risk stratification in hypertrophic cardiomyopathy using machine learning wall thickness measurement: A comparison with human test-retest performance. *Lancet Digit. Health* **2021**, *3*, e20–e28. [[CrossRef](#)]
37. Tobon-Gomez, C.; Butakoff, C.; Yushkevich, P.; Hugué, M.; Frangi, A.F. 3D Mesh Based Wall Thickness Measurement: Identification of Left Ventricular Hypertrophy Phenotypes. In Proceedings of the 2nd Annual International Conference of the IEEE EMBS, Buenos Aires, Argentina, 31 August–4 September 2010. [[CrossRef](#)]
38. Milan, A.; Caserta, M.A.; Avenatti, E.; Abram, S.; Veglio, F. Anti-hypertensive drugs and left ventricular hypertrophy: A clinical update. *Intern. Emerg. Med.* **2010**, *5*, 469–479. [[CrossRef](#)] [[PubMed](#)]
39. Budai, A.; Suhai, F.I.; Csorba, K.; Toth, A.; Szabo, L.; Vago, H.; Merkely, B. Fully automatic segmentation of right and left ventricle on short-axis cardiac MRI images. *Comput. Med. Imaging Graph.* **2020**, *85*, 101786. [[CrossRef](#)]
40. Loshchilov, I.; Hutter, F. Decoupled Weight Decay Regularization. *arXiv* **2019**, arXiv:1711.05101.

41. Lin, T.; Goyal, P.; Girshick, R.B.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. *arXiv* **2017**, arXiv:1708.02002.
42. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv* **2015**, arXiv:1502.03167.
43. Islam, M.M.; Beverung, S.; Steward, R., Jr. Bio-Inspired Microdevices that Mimic the Human Vasculature. *Micromachines* **2017**, *8*, 299. [[CrossRef](#)]
44. de Hartog-Keyzer, J.M.L.; El Messaoudi, S.; Harskamp, R.; Vart, P.; Ringoir, L.; Pop, V.; Nijveldt, R. Electrocardiography for the detection of left ventricular hypertrophy in an elderly population with long-standing hypertension in primary care: A secondary analysis of the CHELLO cohort study. *Cardiovasc. Ultrasound* **2020**, *10*, e038824. [[CrossRef](#)]