

Article

Deep Learning Supplants Visual Analysis by Experienced Operators for the Diagnosis of Cardiac Amyloidosis by Cine-CMR

Philippe Germain ^{1,*}, Armine Vardazaryan ^{2,3}, Nicolas Padoy ^{2,3}, Aissam Labani ¹, Catherine Roy ¹, Thomas Hellmut Schindler ⁴ and Soraya El Ghannudi ^{1,5}

¹ Department of Radiology, Nouvel Hopital Civil, University Hospital, 67000 Strasbourg, France; aissam.labani@chru-strasbourg.fr (A.L.); catherine.roy@chru-strasbourg.fr (C.R.); soraya.elghannudi-abdo@chru-strasbourg.fr (S.E.G.)

² ICube, University of Strasbourg, CNRS, 67000 Strasbourg, France; vardazaryan@unistra.fr (A.V.); npadoy@unistra.fr (N.P.)

³ IHU (Institut Hopitalo-Universitaire), 67000 Strasbourg, France

⁴ Mallinckrodt Institute of Radiology, Division of Nuclear Medicine, Washington University School of Medicine, Saint Louis, MO 63110, USA; thschindler@wustl.edu

⁵ Department of Nuclear Medicine, Nouvel Hopital Civil, University Hospital, 67000 Strasbourg, France

* Correspondence: germain.philippe7@gmail.com



Citation: Germain, P.; Vardazaryan, A.; Padoy, N.; Labani, A.; Roy, C.; Schindler, T.H.; El Ghannudi, S. Deep Learning Supplants Visual Analysis by Experienced Operators for the Diagnosis of Cardiac Amyloidosis by Cine-CMR. *Diagnostics* **2022**, *12*, 69. <https://doi.org/10.3390/diagnostics12010069>

Academic Editors: Sameer Antani and Sivaramkrishnan Rajaraman

Received: 9 December 2021

Accepted: 27 December 2021

Published: 29 December 2021

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



Copyright: © 2021 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/>).

Abstract: Background: Diagnosing cardiac amyloidosis (CA) from cine-CMR (cardiac magnetic resonance) alone is not reliable. In this study, we tested if a convolutional neural network (CNN) could outperform the visual diagnosis of experienced operators. Method: 119 patients with cardiac amyloidosis and 122 patients with left ventricular hypertrophy (LVH) of other origins were retrospectively selected. Diastolic and systolic cine-CMR images were preprocessed and labeled. A dual-input visual geometry group (VGG) model was used for binary image classification. All images belonging to the same patient were distributed in the same set. Accuracy and area under the curve (AUC) were calculated per frame and per patient from a 40% held-out test set. Results were compared to a visual analysis assessed by three experienced operators. Results: frame-based comparisons between humans and a CNN provided an accuracy of 0.605 vs. 0.746 ($p < 0.0008$) and an AUC of 0.630 vs. 0.824 ($p < 0.0001$). Patient-based comparisons provided an accuracy of 0.660 vs. 0.825 ($p < 0.008$) and an AUC of 0.727 vs. 0.895 ($p < 0.002$). Conclusion: based on cine-CMR images alone, a CNN is able to discriminate cardiac amyloidosis from LVH of other origins better than experienced human operators (15 to 20 points more in absolute value for accuracy and AUC), demonstrating a unique capability to identify what the eyes cannot see through classical radiological analysis.

Keywords: cardiac amyloidosis; AL/TTR amyloidosis; hypertrophic cardiomyopathy; left ventricular hypertrophy; deep learning; convolutional neural network

1. Introduction

Cardiac amyloidosis (CA) is a specific cardiomyopathy caused by the deposition of misfolded amyloid fibrils in the extracellular myocardial space. Light-chain (AL) and transthyretin (TTR) are the most common subtypes. Cardiac amyloidosis is a fatal disease requiring rapid diagnosis for patients to benefit from recently released medications [1–3]. Its diagnosis has gained significant improvements in recent years, in particular with the recognition of diphosphonate SPECT imaging for the identification of the TTR form of the disease [4].

MRI plays an important role in this field thanks to gadolinium injections providing quite a specific pattern of myocardial late-enhancement [5] and demonstrating highly relevant extracellular volume (ECV) increase [6]. Despite recent relief in the restrictions

on the use of gadolinium chelates [7], caution needs to be exercised in case of renal impairment, and a diagnostic approach without injection would be beneficial. Steady-state free precession (SSFP) cine-CMR is a basic method in cardiac MRI, offering a good quality morphological and functional depiction of important cardiac features [8]. Myocardial wall thickening, atrial enlargement and pericardial or pleural effusion constitute the hallmarks of amyloid cardiac involvement [9]. However, these signs are very nonspecific since they are also seen in many other etiologies of left ventricular hypertrophy such as advanced hypertensive disease, aortic stenosis and other overload diseases such as Fabry disease and sarcomeric hypertrophic cardiomyopathies, which is why cine-CMR alone is not recognized as effective for diagnosing cardiac amyloidosis.

Machine learning and, particularly, deep learning applied to imaging quickly established themselves in most pathological areas, and these methods are now recognized as having diagnostic capacities similar to experienced radiologists, particularly in cardiomyopathies [10] and cardiac amyloidosis [11]. An even more interesting fact concerns the superior diagnostic capacities of deep learning over human readers in some fields, such as breast cancer [12], especially its ability to identify pathologies invisible to the naked eye, such as abnormalities discernible only in immunohistochemistry or through genetic analysis. For example, deep learning was reported to be efficient in improving mutation prediction in hypertrophic cardiomyopathy using MR-cine images [13].

This innovative concept led us to initiate the present study in which we compared the performance of commonly available deep learning methods to experienced radiologists to discriminate cardiac amyloidosis from other myocardial hypertrophies based on cine-CMR alone. Moreover, we explored the capacity of deep learning to differentiate AL from TTR amyloidosis, which is not reliably achievable visually with cine-CMR.

2. Materials and Methods

2.1. Study Population

We retrospectively analyzed the cine-CMR sequences of patients performed between 2010 and 2020 in the radiology department of our hospital. This study was registered and approved by the Institutional Review Board of our university hospital, and all datasets were obtained and de-identified, with waived consent in compliance with the rules of our institution. The cine-CMR exams of 241 patients were studied, including 119 with histologically proven amyloidosis and 122 with left ventricular hypertrophy without amyloidosis (LVH). The patients' characteristics are listed in Table 1.

The left ventricular hypertrophy without amyloidosis group ($n = 122$) consisted of patients referred to CMR for suspected cardiac amyloidosis due to several suggestive features such as a heart failure episode, thickening of the myocardial walls on ultrasound examination, restrictive transmitral Doppler filling pattern, reduced longitudinal strain with apical sparing, monoclonal gammopathy or dubious Perugini grade 1 bone scintigraphy. Other cases presented a CMR of concentric left ventricular hypertrophy (left ventricular wall thickness ≥ 13 mm in diastole). The clinical context was consistent with hypertension, aortic stenosis or non-obstructive hypertrophic cardiomyopathy. Late-enhancement imaging obtained in all cases never demonstrated circumferential subendocardial or diffuse late-enhancement patterns suggestive of amyloid involvement.

For the amyloidosis group ($n = 119$), the selection criteria for amyloidosis diagnosis were based on typical CMR features confirmed by clinical, biological, bone scintigraphic and anatomic-histological findings. Left ventricular wall thickening (≥ 13 mm in diastole), left \pm right atrial dilatation, increased native myocardial T1 relaxation time and/or extracellular volume (ECV), pericardial or pleural effusion and typical subendocardial late-enhancement pattern (circumferential, diffuse or not related to a coronary territory) were the main diagnostic clues for amyloidosis.

Table 1. Clinical and CMR characteristics of the study population.

	Amyloidosis	LVH	<i>p</i>
N patients	119	122	
Age (years)	74.65 ± 9.53	59.50 ± 14.34	0.0001
Sex (F/M)	31/88	39/83	0.31
Weight (kg)	70.80 ± 15.16	82.95 ± 20.50	0.0001
Height (m)	169.9 ± 8.84	170.36 ± 10.05	0.78
BSA (m ²)	1.84 ± 0.22	2.00 ± 0.27	<0.0001
IVS (mm)	18.11 ± 3.54	18.38 ± 3.54	0.56
LVMI (g/m ²)	115.96 ± 29.08	116.58 ± 31.43	0.88
LVDVI (mL/m ²)	69.88 ± 22.21	74.51 ± 20.82	0.36
LVEF (%)	58.96 ± 10.93	67.33 ± 12.18	<0.0001
LA surface (cm ²)	31.55 ± 5.23	25.47 ± 5.96	0.0002
Systolic time (ms)	321 ± 39	332 ± 40	0.095
T1 (ms)	1138.5 ± 48.1	1038.0 ± 56.2	<0.0001
ECV (%)	53.97 ± 11.17	26.89 ± 4.00	<0.0001
N long axis frames/patient	2.24 ± 0.93	2.22 ± 0.94	0.93
N short axis frames/patient	3.41 ± 1.45	3.59 ± 1.27	0.49
N frames/patient	5.68 ± 1.85	5.47 ± 1.81	0.58
N frame post-gadolinium	171/676	167/667	0.96
N patient with pericard.	54 (45%)	27 (22%)	0.00013
N patients with pleural.	45 (38%)	10 (8%)	0.00001
N patients with both.	24 (20%)	3 (2.5%)	0.00001

The characteristics of patients with amyloidosis and left ventricular hypertrophy were included in this study. The number of observations, (integer) or average values ± standard deviation, are listed: BSA: body surface area; IVS: interventricular septum thickness; LVMI: left ventricular mass index; LVDVI: left ventricular diastolic volume index; LVEF: left ventricular ejection fraction; LA: left atrial; systolic time: the time of the systolic image; and ECV: extracellular volume. Between the parentheses is the percentage. Pericard. is for pericardial effusion, pleural is for pleural effusion and both is for pericardial + pleural effusions.

The characteristics of AL and TTR patients can be found in the supplemental material (Table S1). TTR amyloidosis was defined in 38 patients without monoclonal gammopathy and with a ^{99m}Tc-diphosphonate SPECT Perugini score of >1 or with amyloid deposits on an extracardiac and/or endomyocardial biopsy. AL amyloidosis was reported in 59 cases, based on the detection of a kappa/lambda free light-chain with monoclonal gammopathy and an extracardiac and/or endomyocardial biopsy. Among the 22 patients who were not categorized as AL or TTR, three were AA type, three had uncertain immunostaining, one had Perugini 1 and no gammopathy, four elderly patients died and 11 were lost to follow-up.

For the cine-CMR acquisitions, all images were obtained at 1.5 Tesla, using three Siemens (Erlangen, Germany) and one Philips (Eindhoven, The Netherlands) scanners. Steady-state free precession (SSFP) cine sequences were obtained with TE/TR 1.6/3.5 ms, 8 to 32 elements cardiac coil and 6 to 8 mm thick slices. End-systole (with the smallest left ventricular dimension) was visually selected (systolic time in Table 1). Orientation planes were long axis (4-chamber and vertical 2-chamber views) and short axis views. Table 1 lists the summary of acquisition parameters.

2.2. Image Preparation

The image preparation of cine studies exported from the PACS of our hospital was carried out with a dedicated Visual C software. All images were first de-identified and resampled (bilinear) in order to obtain a normalized homogeneous pixel size of 1.5 mm. The images' intensity windowing was manually focused on the central cardiac region of interest. Diastolic and systolic frames were selected. Epicardial contours (ROI_epi) and myocardial contours (ROI_myo) were manually drawn.

Finally, five pairs of images (cropped to 128 and 160 pixels, full view 256 pixels, ROI_epi and ROI_endo), as illustrated in Figure 1, were stored. The purpose of these tests (especially for ROIs) was to determine if a focused analysis led to better classification performance. Labeling (orientation plane, pathology, presence of effusion and gadolinium injection) was carried out simultaneously and saved in the labeled file.

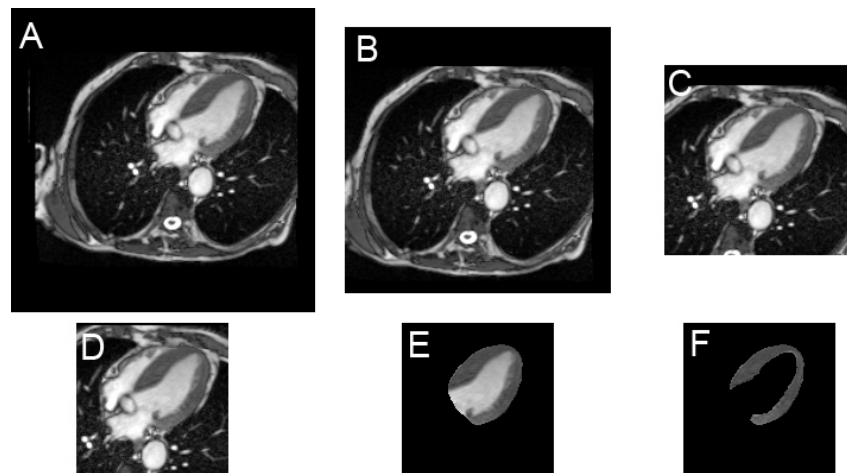


Figure 1. Example of input shapes submitted to the CNN, with native 256×256 full image format (A), 224×224 cropped image (B), 160×160 cropped image (C), 128×128 cropped image (D), epicardial region of interest (ROI) image (E) and myocardial ROI (F).

2.3. Deep Learning Process

CNN implementation was performed in Python 3.7.6, with Keras library and TensorFlow backend. According to CLAIM recommendations [14], the data were distributed in order to ensure that images of the same patient always lie in either the train set, the validation set or the test set (no mixture between these sets).

For hyperparameter trimming, data processing was performed according to the diagram shown in Figure 2. A 40% test set (538 pairs of frames and 96 patients) was isolated and stored as a held-out test set. With the 60% remaining data, a three-fold cross-validation training was performed in order to trim hyperparameters (batch size, optimizer, learning rate, decay, number of trainable layers, dropout rate and parameters of the image data generator). This was done to avoid the influence of individual training and validation examples on the choice of hyperparameters.

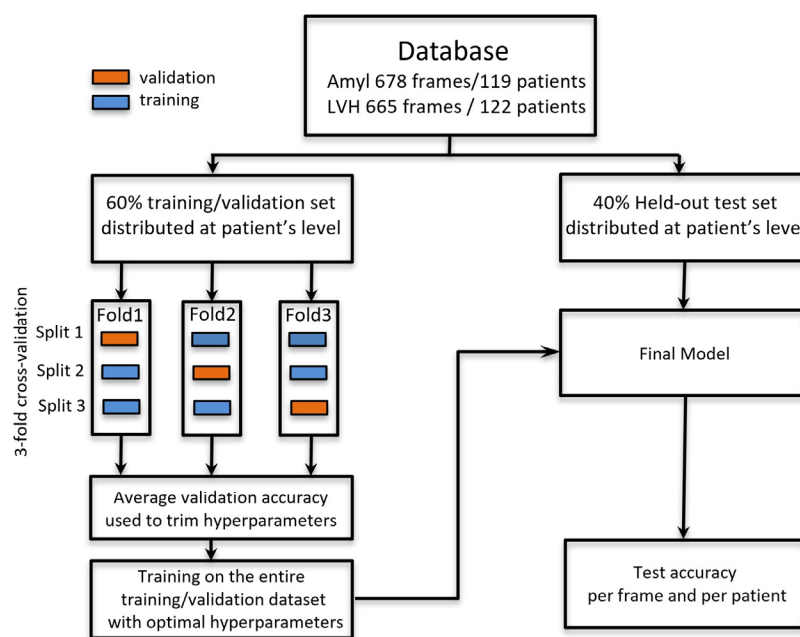


Figure 2. Schematic view of the processing method used in order to strictly separate training/validation data and test data.

With optimal hyperparameters, a final model was built with all training data and evaluated on the test set. Patient-based metrics were calculated from the average of the predicted probability corresponding to all frames of a unique patient.

A VGG16 [15] base model was used and trained from scratch for diastolic and systolic frames. The two outputs (diastole and systole) were concatenated and followed by the following layers, where ReLU non-linearity was used after each Dense layer except the last one: Flatten, Dense 256, Dropout 0.40, Dense 128, Dropout 0.45, Dense 64, Dropout 0.50, Dense 1 and output Sigmoid activation layer. In the final model, training was done with batch size: 32; number of epochs: 150; optimizer: SGD; LR 4×10^{-5} ; and decay: 10^{-6} . Binary cross entropy was used as a loss function. The parameters of data augmentation applied during training were zoom range <0.15 , 15% height and width shift range and up to 20° rotation.

The Grad-CAM algorithm [16] was used to visualize class activation maps. With this algorithm, the identification of the most contributive pixels involved for each class is related to the gradient information flowing into the final convolutional layer of the network.

2.4. Experienced Radiologists/Cardiologists Blind Reading

The blind reading of diastolic and systolic images was performed by one radiologist and two cardiologists (>10 years' experience of CMR analysis and reporting). Frame-based reading was obtained from the pairs of images corresponding to the test set. Patient-based reading was obtained from the whole dataset (241 patients), and paired comparisons were made with the 40% held-out test set (average number of frame pairs, 5.5 per patient).

2.5. Evaluation and Statistical Analysis

The performance metrics—computed on a frame-basis and a patient-basis—were test accuracy, sensitivity, specificity, confusion matrices, receiver operating characteristic (ROC) curves and precision-recall curves with the corresponding area under the curve (AUC) values. Testing the relationship between categorical variables (e.g., accuracy comparisons) was carried out with a Chi-square test. A comparison of the quantitative values was performed with Student's *t*-test, and a comparison of the AUC of ROC curves was performed with the Delong test. MedCalc 12.1.4 (MedCalc Software, Ostend, Belgium) was used for statistical analyses.

3. Results

3.1. Amyloidosis vs. LVH Classification Obtained with the Held-Out Test Set According to the Input Shape

Table 2 lists the results obtained with the various input shapes illustrated in Figure 1. Patient-based results were always better than frame-based results.

Table 2. Accuracy and AUC of the ROC curve for classification of amyloidosis vs. LVH in the 40% held-out test group, according to the input shape.

Input Shape	Frame-Based		Patient-Based	
	Accuracy	ROC AUC	Accuracy	ROC AUC
160 × 160/D + S	0.759	0.836 [0.786–0.878]	0.812	0.937 [0.828–0.987]
160 × 160/D	0.760 (ns)	0.820 (ns) [0.769–0.864]	0.833 (ns)	0.918 (ns) [0.802–0.978]
160 × 160/S	0.733 (ns)	0.801 (0.04) [0.749–0.848]	0.833 (ns)	0.890 (ns) [0.767–0.962]
256 × 256/D + S	0.710 (ns)	0.790 (0.03) [0.735–0.836]	0.771 (ns)	0.803 (0.02) [0.663–0.904]
224 × 224/D + S	0.728 (ns)	0.823 (ns) [0.772–0.867]	0.812 (ns)	0.852 (ns) [0.720–0.938]
128 × 128/D + S	0.740 (ns)	0.808 (ns) [0.756–0.853]	0.812 (ns)	0.922 (ns) [0.807–0.979]
Epicardial ROI	0.722 (ns)	0.787 (0.01) [0.762–0.810]	0.791 (ns)	0.888 (ns) [0.839–0.927]
Myocard. ROI	0.662 (0.05)	0.719 (0.01) [0.693–0.745]	0.714 (ns)	0.814 (0.03) [0.756–0.863]

Results obtained with the 40% held-out test set after hyperparameters tuning. 160 × 160 indicates the cropping size of input frames. D and S indicate diastole and systole. Between brackets is the confidence interval of AUC. Values between parentheses indicate the level of significance of the difference as compared to the 160 × 160 D + S result (assessed with Chi-square test from the number of observations for accuracy and assessed by Delong test for AUC comparisons).

Optimal performance was obtained with 160 × 160 cropped diastolic and systolic images in which per frame analysis provided a test accuracy of 0.759 and an AUC of 0.836, whereas per patient analysis provided a test accuracy of 0.812 and an AUC of 0.937.

Combining diastole and systole did not improve the results. Full field 256 × 256 frames and focused myocardial ROI images provided significantly weaker results.

3.2. Amyloidosis vs. LVH Classification Obtained with the Held-Out Test Set by Human Readers and by CNN

The comparison between classification by experienced radiologists/cardiologists and the CNN is given in Table 3. The CNN provided a largely superior performance when compared to human readers.

Frame-based comparisons of human vs. CNN classification led to an accuracy of 0.605 vs. 0.746 ($p < 0.0008$) and an AUC of 0.630 vs. 0.824 ($p < 0.0001$).

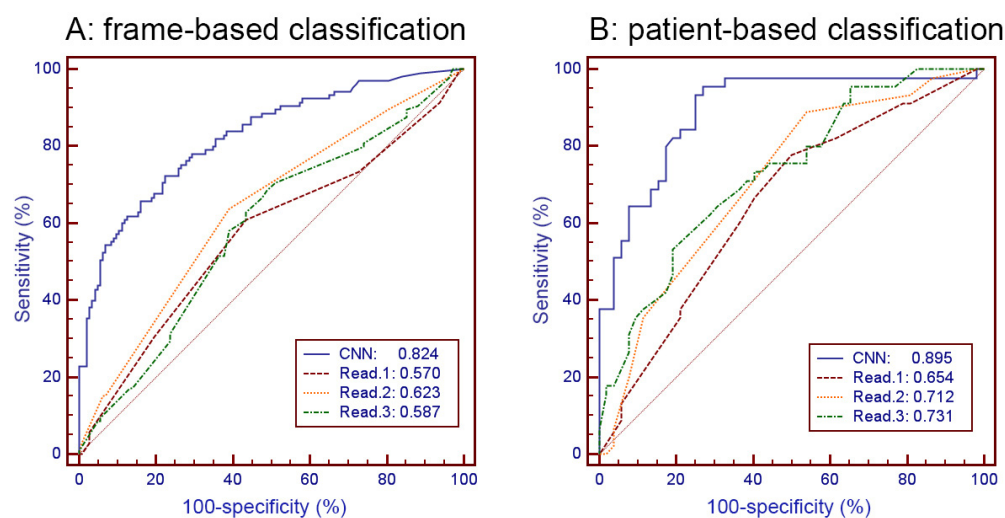
Patient-based comparisons provided an accuracy of 0.660 vs. 0.825 (0.008) and an AUC of 0.727 vs. 0.895 ($p < 0.002$). The ROC curves of these comparisons are plotted in Figure 3.

Table 3. Accuracy and AUC of the ROC curve for classification of amyloidosis vs. LVH in the held-out test group for human readers vs. CNN.

Metric	Frame-Based			Patient-Based		
	Accur.	Sensitiv. Specific.	ROC AUC	Accur.	Sensitiv. Specific.	ROC AUC
CNN	0.746	77.0 71.0	0.824 [0.770–0.869]	0.825	85.7 77.6	0.895 [0.816–0.948]
Read 1	0.585 (0.001)	66.4 50.85	0.570 [0.506–0.632] (0.0001)	0.629 (0.004)	67.4 58.8	0.654 [0.550–0.747] (0.001)
Read 2	0.623 (0.005)	69.6 54.5	0.623 [0.560–0.684] (0.0001)	0.649 (0.009)	69.6 60.8	0.712 [0.611–0.799] (0.0002)
Read 3	0.585 (0.001)	66.4 50.9	0.587 [0.523–0.649] (0.0001)	0.660 (0.013)	71.1 61.5	0.731 [0.631–0.816] (0.002)
Read (avg)	0.605 (0.0008)	69.2 52.7	0.630 [0.567–0.690] (0.0001)	0.660 (0.008)	72.1 61.1	0.727 [0.627–0.813] (0.002)

Frame-based and patient-based results obtained with the held-out test set by human readers and by CNN. Accur. is for accuracy, Sensitiv. and Specific. are for sensitivity and specificity. Values between parentheses indicate the level of significance of the difference between human reader and CNN (assessed with Chi-square test from the number of observations for accuracy and assessed by Delong test for AUC comparisons).

Amyloidosis vs LVH

**Figure 3.** ROC curves and AUC for frame-based (A) and patient-based (B) classification of amyloidosis vs. LVH by CNN and by three human readers (Read. 1 to 3).

3.3. CNN Classification of AL vs. TTR Amyloidosis

The frame-based accuracy and AUC obtained by the CNN classification of AL vs. TTR cardiac amyloidosis were 0.662 and 0.703 [0.664–0.741]. The corresponding patient-based values were 0.711 and 0.752 [0.654–0.834]. No comparison was performed here with human classification, but the comparison between the AUC values of the CNN and the simple left ventricular septal wall thickness measurement (per-patient AUC 0.735) did not show a statistically significant difference.

3.4. Analysis of the Saliency Maps

Saliency maps, which reveal the pixel areas responsible for classification, show that cardiac regions contribute to CNN decisions in only 25% of cases (Figure 4). Among the extracardiac targeted regions, the lungs are the most frequent, followed by the subcutaneous fat and liver. Distribution is quite similar for correct classification (concordant) and erroneous classification (discordant).

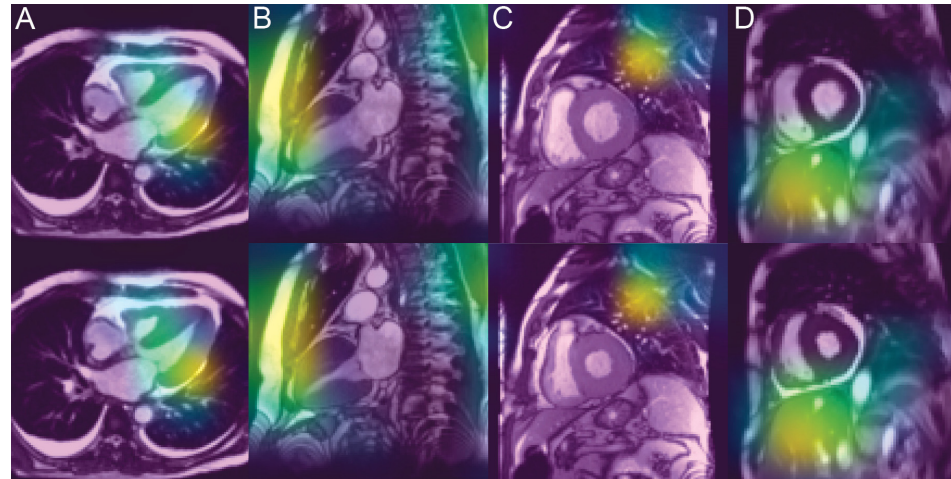


Figure 4. Saliency maps targeting cardiac region (A) but also frequently subcutaneous fat (B), lung (C) or liver (D). Diastolic frames are shown in the upper row and systolic frames in the lower row.

4. Discussion

The most important result of this study is the possibility of discriminating cardiac amyloidosis and LVH from other origins by simple cine-CMR images, which was significantly better with the CNN than by the physicians' visual analysis. The comparison carried out on slightly more than 100 patients of the two groups shows that, for frame-based and patient-based analysis, binary classification accuracy is approximately 15 absolute points higher with the CNN than with experienced radiologists/cardiologists. The same significant difference is also found by considering the AUC of the ROC curve, with a little less than 20 points absolute value improvement with the CNN as compared with experienced human readers.

4.1. Methodological Considerations

Patient-based analysis constitutes a much more relevant assessment because this is how the clinical diagnosis is carried out, and it should be noted that the transposition of the results from the image-level to the patient-level (by taking the average of the elementary predictions per frame) leads to an improvement of the accuracy in the range of 5 points (absolute value) and in the range of 10 points for the AUC. This phenomenon, which is observed for the human reader and CNN, may be explained thanks to the "averaging process" in the mind of the physician who examines the whole set of pictures of the patient.

The influence of methodological choices should be stressed: (1) The distribution of patients' images in a distinct train or validation/test data sets is mandatory; otherwise, the results would be clearly biased because we would have trained on images that are—for some features—similar to test images. Processing this way (without frame distribution for a unique patient) with our data set led to a misleading "improvement" of almost 10 points (absolute value) for accuracy and AUC results (data not listed here). (2) The strict separation of the train and validation set for hyperparameter tuning and the test set has been done. This method, based on a separate test set, schematized in Figure 2, is required to avoid information leakage related to hyperparameter tuning.

4.2. Superiority of CNN Capacities over Human Diagnosis

The aim of this study was not to propose making the diagnosis of cardiac amyloidosis solely on the cine-CMR data because much more relevant CMR indices are available thanks to gadolinium injection. Actually, late-enhancement and ECV allow the diagnosis of the presence of CA with a high sensitivity of 95% and an even higher specificity of 98% [5], and deep learning was demonstrated to be efficient in this field [11]. Our goal was to show that deep learning is able to extract diagnostic clues clearly surpassing visual analysis (15 to 20 points in the present study).

Excellent performances of the CNN are often reported in the literature, but their interest is limited if they are not compared to human performance. Among human-machine comparisons, many studies have reported that CNN diagnosis is on par with human visual assessment in multiple areas [17]. For example, for malignancy risk estimation of pulmonary nodules using thoracic CT, Venkadesh et al. [18] reported that the DL algorithm had an AUC of 0.96, which was significantly better than the average AUC of the clinicians (0.90) but comparable to that of thoracic radiologists. Our model was able to discriminate between AL and TTR CA with interesting values of patient-based accuracy (0.711) and AUC (0.752); however, this was no better than the classification obtained with the simple measurement of the septal thickness, already reported in previous publications [19–21] and resulting from the known increased amyloid burden in this subtype.

Of more interest is to show significant machine-over-human superiority in routine areas, where “clinical” visual analysis is the classic benchmark. Our study provides an interesting demonstration in this direction for diagnosing cardiac amyloidosis from cine-CMR. A small number of other publications could demonstrate that AI systems are capable of surpassing human experts in disease prediction. Such is the case for the distinction between low-grade and high-grade glioma by radiologists, which lacks accuracy (40–45% of non-enhancing MR lesions are found subsequently to be malignant glioma), whereas, in contrast, CNN-based grading provides > 90% accuracy [22]. Resnet-50 CNN outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task [23]. For the diagnosis of breast cancer, in a large multicenter study, Mc Kinney et al. [12] found that the AI system exhibited specificity and sensitivity superior to that of radiologists practicing in an academic medical center and exceeded the average performance of radiologists by a significant improvement in the area under the ROC curve ($\Delta\text{AUC} = +0.115$). Similarly, in differentiating benign from malignant renal tumors, Xu et al. [24] reported higher AUC with the CNN model (0.906, based on T2-weighted images) as compared to the AUC obtained by two radiologists (0.724).

4.3. Unveiling the Invisible

One more step in this diagnostic quest is the possibility of discriminating pathological conditions that clinicians are not able to predict at all using the naked eye. Subtyping molecular markers, histological or immune-histochemical and genetic classes is impossible to ascertain from radiologic data. These identifications were initially proposed from radiomic signatures, for instance, to discriminate between hypertensive heart disease and hypertrophic cardiomyopathy [25] or between recent infarction vs. old infarction [26]. However, several comparative studies have demonstrated that deep learning based on radiologic data is superior to radiomics. This has been demonstrated for renal cancer [24], subtyping different types of cerebral glioma [27], diagnosis of breast cancer [28] and predicting axillary lymph node metastasis of breast cancer [29].

This may be explained because radiomics’ features are handcrafted in advance and, thus, may not always fit to discriminate particular tasks. In contrast, the CNN is more flexible, adaptive and dynamic. As a data-driven tool, it is able to automatically learn to extract and select task-specific features if the amount of training data is large enough. Further evidence for the power of deep learning to make a histological diagnosis from radiological data has been provided by Zhao et al. for renal cell carcinomas Fuhrman-grading [30] and by Yuan et al. for prostate cancer Gleason score staging (accuracy 0.87) [31].

4.4. Explanation of Classification Remains Unsatisfactory

Deep neural networks operate through a multilayer nonlinear structure, making their predictions difficult to interpret. They are able to pick up a number of features that cannot be interpreted by humans but which are relevant for making a diagnosis. These automatically-learned discriminative features are unfortunately presently not clearly identifiable.

Grad-CAM helps identify the areas of pixels that are most responsible for class prediction [16]. This should provide valuable clues to understand the algorithm's decision. In principle, the salient areas should be located in the cardiac region, which only appeared in a quarter of the cases in our study. Two explanations may be advanced for this anomaly.

(1) Technically, our network uses only fully connected layers in the last phase, which is where the classification happens, but saliency cannot be obtained from fully connected layers. As a solution, we should try replacing some of the fully connected layers that come right after VGG with convolution. This way, the spatial information would be preserved longer in the network, and we might see more meaning in the saliency maps.

(2) Amyloidosis is not a disease confined to the heart since the involvement of the lungs, fatty tissues and other organs is also common. Liver and, moreover, spleen amyloid deposits have been reported in 41% of patients with systemic amyloidosis (almost only in AL type), and CMR-derived ECV measurement showed good diagnostic capability in this field [32]. This is why the diagnosis is also based on extracardiac biopsies, and it is interesting to note that the texture analysis was able to show specificities in the architecture of ultrasound images within abdominal fat [33], resulting in increased echogenicity and a loss of the normal structure of the fat layer, consistent with histopathological amyloid deposition in the fat.

This ubiquitous aspect of the disease may also explain why the input shape submitted to the CNN (from the full field image to the small region of interest focused on the sole myocardium illustrated in Figure 1) hardly modifies the performance of our model as shown in Table 2. It can also be noted in Table 2 that the combination of diastole and systole does not provide any diagnostic benefit, unlike for other cardiomyopathies [10], because the global LV systolic function is generally preserved in the early stage of amyloidosis.

4.5. Study Limitations

Two types of confounding factors must be mentioned. First, plane orientation and the presence of gadolinium in the sets of images could have influenced the results, but Table 1 shows a perfect equivalence between the two groups. Second, the presence of pericardial or pleural effusion constitutes a more important bias because the prevalence (slightly higher than in the study of Binder et al. [9]) is very different in the two groups. Pericardial effusion is observed in almost 50% of CA, i.e., two times more often than in hypertrophies unrelated to amyloidosis. Pleural effusions are observed in just over a third of CA, i.e., four times more than in other hypertrophies, and mixed effusions are 10 times more frequent in the amyloidosis group than in the LVH group. This disparity probably contributes to the classification made by CNNs (although heat maps rarely focus on areas of effusion) but also influences clinical judgment, so that the bias is the same for the machine and for the human, which, therefore, does not explain the diagnostic superiority of the algorithm.

A multiparametric approach is needed. Only cine-CMR data has been used here, and it is likely that one could significantly improve performances by combining the analysis with other CMR sequences such as T1 mapping, ECV assessment and late gadolinium-enhancement imaging. Based on gadolinium-enhanced images—and not on cine-MR images—Martini et al. obtained an accuracy of 0.88 and AUC of 0.98 [11], but remember that our aim was not to develop the best model to optimize cardiac amyloidosis diagnosis but to compare CNN and human reader performance. For the distinction between AL and TTR CA, it has been reported that transmural patterns of late gadolinium enhancement may differentiate these two types of the disease [21] but with relatively low performance. Recently, the use of a logistic regression model integrating T2 mapping (slightly increased

in the AL subtype) and right ventricular ejection fraction combined with age was reported to discriminate between these two subtypes with an AUC of 0.92 [34]. The performance of AI integrating such multiparametric CMR features, especially for the distinction between AL and TTR cardiac amyloidosis, should be explored in the future.

Technical improvements should be implemented. The leverage of more sophisticated CNN models (not limited to the classical VGG model used here) and, moreover, the combination (concatenation) of several multiparametric inputs, with possible additional categorical clinical input variables (e.g., [30]), should improve performance. Orientation plane specific models [11] should also be tested since images of different views were classified here by the same network, which makes learning relevant features from images potentially much harder as it increases variability unrelated to any disease. Significant work also remains to be done to improve the explainability of the results. Finally, the relatively limited number of observations and the monocentric nature of this study constitute another limitation. Multicenter studies could be of interest for the further validation and generalization of our findings.

5. Conclusions

In this study, based on cine-CMR images alone, we could demonstrate the ability of CNNs to discriminate cardiac amyloidosis from LVH of other origins significantly better than experienced human operators. The diagnostic accuracy and AUC were 15 to 20 points higher (in absolute value) for the VGG convolutional network used here than for human readers. This diagnostic superiority of the CNN results from the unique capability of the algorithm to identify features invisible to the naked eye, indiscernible through the classical radiological analysis. This scientific novelty, already reported in a few recent articles concerning other pathological fields, opens up promising prospects for improving diagnostic capacities in routine clinical practice. The astonishing potential of CNNs to improve the recognition of pathologies that are imperfectly detectable in radiology and to reveal invisible clues such as the histological type of lesions will certainly constitute a large field of future research.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/diagnostics12010069/s1>; Table S1: Clinical and CMR characteristics of AL and TTR cardiac amyloidosis.

Author Contributions: Conceptualization, P.G. and S.E.G.; methodology, N.P. and P.G.; software, A.V.; validation, P.G., N.P. and S.E.G.; formal analysis, A.L., A.V. and T.H.S.; investigation, S.E.G.; data curation, P.G. and S.E.G.; writing—original draft preparation, P.G.; writing—review and editing, S.E.G., A.L., N.P., C.R. and T.H.S.; supervision, N.P. and S.E.G.; project administration, C.R. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by French state funds managed by the ANR under reference ANR-10-IAHU-02, without any involvement in the study design, data gathering, analysis/interpretation of data or writing of the report.

Institutional Review Board Statement: This retrospective study was registered and approved by the Institutional Review Board of the university hospital of Strasbourg (ref 20–072/sept 2020). All datasets were obtained and de-identified, with waived consent in compliance with the Institutional Review Board of our institution.

Informed Consent Statement: All datasets were obtained and de-identified, with waived consent in compliance with the Institutional Review Board of our institution. No protected health information for any subject is given in this manuscript.

Data Availability Statement: The database and code can be made available by reasonable request after the agreement of the Clinical Research Department of our hospital.

Conflicts of Interest: Nicolas Padoy serves as a consultant for Caresyntax and has received research support from Intuitive Surgical, unrelated to this work. Soraya El Ghannudi serves as a consultant for Pfizer.

References

1. Garcia-Pavia, P.; Rapezzi, C.; Adler, Y.; Arad, M.; Basso, C.; Brucato, A.; Burazor, I.; Caforio, A.L.P.; Damy, T.; Eriksson, U.; et al. Diagnosis and treatment of cardiac amyloidosis. A position statement of the European Society of Cardiology Working Group on Myocardial and Pericardial Diseases. *Eur. J. Heart Fail.* **2021**, *23*, 512–526. [[CrossRef](#)] [[PubMed](#)]
2. Kittleson, M.M.; Maurer, M.S.; Ambardekar, A.V.; Bullock-Palmer, R.P.; Chang, P.P.; Eisen, H.J.; Nair, A.P.; Nativi-Nicolau, J.; Ruberg, F.L. American Heart Association Heart Failure and Transplantation Committee of the Council on Clinical Cardiology. Cardiac Amyloidosis: Evolving Diagnosis and Management: A Scientific Statement From the American Heart Association. *Circulation* **2020**, *142*, 7–22. [[CrossRef](#)] [[PubMed](#)]
3. Papathanasiou, M.; Carpinteiro, A.; Rischpler, C.; Hagenacker, T.; Rassaf, T.; Luedike, P. Diagnosing cardiac amyloidosis in every-day practice: A practical guide for the cardiologist. *Int. J. Cardiol. Heart Vasc.* **2020**, *28*, 100519. [[CrossRef](#)]
4. Maurer, M.S.; Bokhari, S.; Damy, T.; Dorbala, S.; Drachman, B.M.; Fontana, M.; Grogan, M.; Kristen, A.V.; Lousada, I.; Nativi-Nicolau, J.; et al. Expert Consensus Recommendations for the Suspicion and Diagnosis of Transthyretin Cardiac Amyloidosis. *Circ. Heart Fail.* **2019**, *12*, e006075. [[CrossRef](#)] [[PubMed](#)]
5. Chantzantonis, G.; Bietenbeck, M.; Elsanhoury, A.; Tschöpe, C.; Pieske, B.; Tauscher, G.; Viethier, J.; Shomanova, Z.; Mahrholdt, H.; Rolf, A.; et al. Diagnostic value of cardiovascular magnetic resonance in comparison to endomyocardial biopsy in cardiac amyloidosis: A multi-centre study. *Clin. Res. Cardiol.* **2021**, *110*, 555–568. [[CrossRef](#)] [[PubMed](#)]
6. Wang, T.K.M.; Brizneda, M.V.; Kwon, D.H.; Popovic, Z.B.; Flamm, S.D.; Hanna, M.; Griffin, B.P.; Xu, B. Reference Ranges, Diagnostic and Prognostic Utility of Native T1 Mapping and Extracellular Volume for Cardiac Amyloidosis: A Meta-Analysis. *J. Magn. Reson. Imaging* **2021**, *53*, 1458–1468. [[CrossRef](#)]
7. Weinreb, J.C.; Rodby, R.A.; Yee, J.; Wang, C.L.; Fine, D.; McDonald, R.J.; Perazella, M.A.; Dillman, J.R.; Davenport, M.S. Use of Intravenous Gadolinium-based Contrast Media in Patients with Kidney Disease: Consensus Statements from the American College of Radiology and the National Kidney Foundation. *Kidney Med.* **2021**, *3*, 142–150. [[CrossRef](#)] [[PubMed](#)]
8. Martinez-Naharro, A.; Treibel, T.A.; Abdel-Gadir, A.; Bulluck, H.; Zumbo, G.; Knight, D.S.; Kotecha, T.; Francis, R.; Hutt, D.F.; Rezk, T.; et al. Magnetic Resonance in Transthyretin Cardiac Amyloidosis. *J. Am. Coll. Cardiol.* **2017**, *70*, 466–477. [[CrossRef](#)] [[PubMed](#)]
9. Binder, C.; Duca, F.; Binder, T.; Retzl, R.; Dachs, T.M.; Seirer, B.; Camuz Ligios, L.; Dusik, F.; Capelle, C.; Qin, H.; et al. Prognostic implications of pericardial and pleural effusion in patients with cardiac amyloidosis. *Clin. Res. Cardiol.* **2021**, *110*, 532–543. [[CrossRef](#)] [[PubMed](#)]
10. Germain, P.; Vardazaryan, A.; Padoy, N.; Labani, A.; Roy, C.; Schindler, T.H.; El Ghannudi, S. Classification of Cardiomyopathies from MR Cine Images Using Convolutional Neural Network with Transfer Learning. *Diagnostics* **2021**, *11*, 1554. [[CrossRef](#)] [[PubMed](#)]
11. Martini, N.; Aimo, A.; Barison, A.; Della Latta, D.; Vergaro, G.; Aquaro, G.D.; Ripoli, A.; Emdin, M.; Chiappino, D. Deep learning to diagnose cardiac amyloidosis from cardiovascular magnetic resonance. *J. Cardiovasc. Magn. Reson.* **2020**, *22*, 84. [[CrossRef](#)] [[PubMed](#)]
12. McKinney, S.M.; Sieniek, M.; Godbole, V.; Godwin, J.; Antropova, N.; Ashrafian, H.; Back, T.; Chesus, M.; Corrado, G.S.; Darzi, A.; et al. International evaluation of an AI system for breast cancer screening. *Nature* **2020**, *577*, 89–94. [[CrossRef](#)]
13. Zhou, H.; Li, L.; Liu, Z.; Zhao, K.; Chen, X.; Lu, M.; Yin, G.; Song, L.; Zhao, S.; Zheng, H.; et al. Deep learning algorithm to improve hypertrophic cardiomyopathy mutation prediction using cardiac cine images. *Eur. Radiol.* **2021**, *31*, 3931–3940. [[CrossRef](#)] [[PubMed](#)]
14. Mongan, J.; Moy, L.; Kahn, C.E., Jr. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers. *Radiol. Artif. Intell.* **2020**, *2*, e200029. [[CrossRef](#)] [[PubMed](#)]
15. Simonyan, K.; Zisserman, A. Very Deep Convolutional Network for Large Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
16. Selvaraju, R.R.; Das, A.; Vedantam, R.; Cogswell, M.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Int. J. Comput. Vis.* **2019**, *128*, 336–359. [[CrossRef](#)]
17. Shen, J.; Zhang, C.J.P.; Jiang, B.; Chen, J.; Song, J.; Liu, Z.; He, Z.; Wong, S.Y.; Fang, P.H.; Ming, W.K. Artificial Intelligence Versus Clinicians in Disease Diagnosis: Systematic Review. *JMIR Med. Inform.* **2019**, *7*, e10010. [[CrossRef](#)]
18. Venkadesh, K.V.; Setio, A.A.A.; Schreuder, A.; Scholten, E.T.; Chung, K.; Wille, M.M.W.; Saghir, Z.; van Ginneken, B.; Prokop, M.; Jacobs, C. Deep Learning for Malignancy Risk Estimation of Pulmonary Nodules Detected at Low-Dose Screening CT. *Radiology* **2021**, *300*, 438–447. [[CrossRef](#)] [[PubMed](#)]
19. Itzhaki Ben Zadok, O.; Vaturi, M.; Vaxman, I.; Iakobishvili, Z.; Rhurman-Shahar, N.; Kornowski, R.; Hamdan, A. Differences in the characteristics and contemporary cardiac outcomes of patients with light-chain versus transthyretin cardiac amyloidosis. *PLoS ONE* **2021**, *16*, e0255487. [[CrossRef](#)] [[PubMed](#)]
20. Quarta, C.C.; Solomon, S.D.; Uraizee, I.; Kruger, J.; Longhi, S.; Ferlito, M.; Gagliardi, C.; Milandri, A.; Rapezzi, C.; Falk, R.H. Left ventricular structure and function in transthyretin-related versus light-chain cardiac amyloidosis. *Circulation* **2014**, *129*, 1840–1849. [[CrossRef](#)]
21. Dungu, J.N.; Valencia, O.; Pinney, J.H.; Gibbs, S.D.; Rowczenio, D.; Gilbertson, J.A.; Lachmann, H.J.; Wechalekar, A.; Gillmore, J.D.; Whelan, C.J.; et al. CMR-based differentiation of AL and ATTR cardiac amyloidosis. *JACC Cardiovasc. Imaging* **2014**, *7*, 133–142. [[CrossRef](#)] [[PubMed](#)]

22. Zlochower, A.; Chow, D.S.; Chang, P.; Khatri, D.; Boockvar, J.A.; Filippi, C.G. Deep Learning AI Applications in the Imaging of Glioma. *Top. Magn. Reason. Imaging* **2020**, *29*, 115–121. [[CrossRef](#)]
23. Brinker, T.J.; Hekler, A.; Enk, A.H.; Klode, J.; Hauschild, A.; Berking, C.; Schilling, B.; Haferkamp, S.; Schadendorf, D.; Holland-Letz, T.; et al. Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. *Eur. J. Cancer* **2019**, *113*, 47–54. [[CrossRef](#)] [[PubMed](#)]
24. Xu, Q.; Zhu, Q.; Liu, H.; Chang, L.; Duan, S.; Dou, W.; Li, S.; Ye, J. Differentiating Benign from Malignant Renal Tumors Using T2- and Diffusion-Weighted Images: A Comparison of Deep Learning and Radiomics Models Versus Assessment from Radiologists. *J. Magn. Reason. Imaging* **2021**. [[CrossRef](#)]
25. Neisius, U.; El-Rewaidy, H.; Nakamori, S.; Rodriguez, J.; Manning, W.J.; Nezafat, R. Radiomic Analysis of Myocardial Native T1 Imaging Discriminates Between Hypertensive Heart Disease and Hypertrophic Cardiomyopathy. *JACC Cardiovasc. Imaging* **2019**, *12*, 1946–1954. [[CrossRef](#)] [[PubMed](#)]
26. Larroza, A.; Materka, A.; López-Lereu, M.P.; Monmeneu, J.V.; Bodí, V.; Moratal, D. Differentiation between acute and chronic myocardial infarction by means of texture analysis of late gadolinium enhancement and cine cardiac magnetic resonance imaging. *Eur. J. Radiol.* **2017**, *92*, 78–83. [[CrossRef](#)] [[PubMed](#)]
27. Li, Y.; Wei, D.; Liu, X.; Fan, X.; Wang, K.; Li, S.; Zhang, Z.; Ma, K.; Qian, T.; Jiang, T.; et al. Molecular subtyping of diffuse gliomas using magnetic resonance imaging: Comparison and correlation between radiomics and deep learning. *Eur. Radiol.* **2021**, *52*. [[CrossRef](#)] [[PubMed](#)]
28. Truhn, D.; Schradang, S.; Haarburger, C.; Schneider, H.; Merhof, D.; Kuhl, C. Radiomic versus Convolutional Neural Networks Analysis for Classification of Contrast-enhancing Lesions at Multiparametric Breast MRI. *Radiology* **2019**, *290*, 290–297. [[CrossRef](#)]
29. Sun, Q.; Lin, X.; Zhao, Y.; Li, L.; Yan, K.; Liang, D.; Sun, D.; Li, Z.C. Deep Learning vs. Radiomics for Predicting Axillary Lymph Node Metastasis of Breast Cancer Using Ultrasound Images: Don't Forget the Peritumoral Region. *Front. Oncol.* **2020**, *10*, 53. [[CrossRef](#)] [[PubMed](#)]
30. Zhao, Y.; Chang, M.; Wang, R.; Xi, I.L.; Chang, K.; Huang, R.Y.; Vallières, M.; Habibollahi, P.; Dagli, M.S.; Palmer, M.; et al. Deep Learning Based on MRI for Differentiation of Low- and High-Grade in Low-Stage Renal Cell Carcinoma. *J. Magn. Reason. Imaging* **2020**, *52*, 1542–1549. [[CrossRef](#)]
31. Yuan, Y.; Qin, W.; Buyyounouski, M.; Ibragimov, B.; Hancock, S.; Han, B.; Xing, L. Prostate cancer classification with multiparametric MRI transfer learning model. *Med. Phys.* **2019**, *46*, 756–765. [[CrossRef](#)]
32. Chacko, L.; Boldrini, M.; Martone, R.; Law, S.; Martinez-Naharro, A.; Hutt, D.F.; Kotecha, T.; Patel, R.K.; Razvi, Y.; Rezk, T.; et al. Cardiac Magnetic Resonance-Derived Extracellular Volume Mapping for the Quantification of Hepatic and Splenic Amyloid. *Circ. Cardiovasc. Imaging* **2021**, *14*, e012506. [[CrossRef](#)] [[PubMed](#)]
33. Misumi, Y.; Ueda, M.; Yamashita, T.; Masuda, T.; Kinoshita, Y.; Tasaki, M.; Nagase, T.; Ando, Y. Novel screening for transthyretin amyloidosis by using fat ultrasonography. *Ann. Neurol.* **2017**, *81*, 604–608. [[CrossRef](#)] [[PubMed](#)]
34. Slivnick, J.A.; Tong, M.S.; Nagaraja, H.N.; Elamin, M.B.; Wallner, A.; O'Brien, A.; Raman, S.V.; Zareba, K.M. Novel Cardiovascular Magnetic Resonance Model to Distinguish Immunoglobulin Light Chain From Transthyretin Cardiac Amyloidosis. *JACC Cardiovasc. Imaging* **2021**, *14*, 302–304. [[CrossRef](#)]