Peer Review File

Article Information: https://dx.doi.org/10.21037/cdt-23-350

For Reviewer A

Comment1: In Introduction section, "we aimed to investigate discriminators and build discrimination models between HHD and HCM using multiparametric CMR findings and radiomics features derived from cine and LGE images". Both the discriminators and the discrimination models utilize Radscore rather than directly using Radiomics features. It is recommended to modify this part.

Reply1: Thank you for your comments. we have made revisions based on your suggestions.

Changes in the text: In the 3rd paragraph of Introduction section, "radiomics features" has been modified as "radiomics score(radscore)"(see Page 6, line 3).

Comment2: In Methods(MR image acquisition) section, "All slice thicknesses were 8 mm for the short axis". So, what is the slice thickness of the long axis images?

Reply2: Thank you for your comments. The slice thickness of the long axis images were 5mm, which has been added in the Methods(MR image acquisition) section.

Changes in the text: We added "5 mm for the long axis"(see Page 8, line 3).

Comment3: In Methods(Discrimination model building) section, "we compared the area under the receiver operator characteristic curve (AUC) of these methods for various MR scanners". What statistical method was used to compare their AUCs?

Reply3: Thank you for your comments. The likelihood ratio test was used to compare the performance of nested models and the Delong test was used to compare the performance of non-nested models. Which has been added in the Methods(Statistical analysis) section.

Changes in the text: We added "The likelihood ratio test was used to compare the performance of nested models and the Delong test was used to compare the performance of non-nested models"(see Page 12, line 6-8).

Comment4: In Methods(Statistical analysis) section, "Quantitative data are expressed as the median and IQR". Why not mean±standard deviation, do they both not conform to a normal distribution? There is no content related to normality testing in the article.

Reply4: Thank you for your reminder. The quantitative data in this study does not conform to a normal distribution according to the Shapiro-Wilk test. Which has been added in the Methods(Statistical analysis) section.

Changes in the text: We added"Because the quantitative data in this study does not conform to a normal distribution according to the Shapiro-Wilk test. They were expressed as the median and IQR, and the Mann–Whitney U tests were used to compare differences between two groups."(see Page 11, line 17-20).

Comment5: In Results(Discrimination model establishment and evaluation) section, "The mLVEDWT, LVEDVi, LVESVi, LVEF, LVH asymmetry, SAM, quantification of LGE, mid-wall LGE". There is no corresponding description of LVEDVi, LVESVi in the Methods

section.
Reply5: Thank you for your reminder. The description of LVEDVi, LVESVi were added in the Methods(Image analysis) section.
Changes in the text: We added"LVEDV index(LVEDVi), and LVESV index(LVESVi)"(see Page 8, line 13).

Comment6: In Equation(2) , the authors used "maximal EDWT", while in the text, "mLVEDWT" is employed. Please ensure consistency.
Reply6: Thank you for your reminder. We modified all the "mLVEDWT" to "maximal EDWT".
Changes in the text: All the "mLVEDWT" in this manuscript has been modified to "maximal EDWT"(see Page 8, line 14, et al).

For Reviewer B
Comment1: I strongly recommend providing the cutoff values for different models in discriminating between the two diseases. This information holds paramount importance for clinical application.

Reply1: Thank you for your reminder. This has been added in the 2nd paragraph of

Results(Discrimination model establishment and evaluation) section.

Changes in the text: We added "The optimal cut-off value of cine radscore, LGE radscore, combined radscore, clinical model and combined model were -0.40, 0.21, -0.44, 87.81 and 72.02, respectively"(see Page 14, line 14-16).

Comment2: In the discussion section on page 15, the author writes "To simplify the process, the cine radscore is recommended for use in conventional clinical work", could you please explain why cine radscore is more recommended for clinical application?
Reply2: Thank you for your comment. We have already removed this section of content. Because it is not possible to make such a recommendation based on the existing evidence.
Changes in the text: We deleted "To simplify the process, the cine radscore is recommended for use in conventional clinical work"(see Page 14, line 10-19).

Comment3: In page 7, "MLVEDWT was manually measured by an experienced observer ." should be"mLVEDWT was manually measured by an experienced observer." Additionally, I do not recommend abbreviating "maximal" as "m," as this abbreviation can easily be mistaken for "mean."
Reply3: Thank you for your reminder. We have made correction, all the "mLVEDWT" were changed to "maximal LVEDWT".
Changes in the text: All the "mLVEDWT" in this manuscript has been modified to "maximal EDWT"(see Page 8, line 14, et al).

Comment4: In the last line of page 7, the author mentions "LVH asymmetry was defined as an LVEDWT > 1.5-fold the wall thickness of the opposing segment", so could you please

clarify whether it refers to any location's wall thickness being greater than 1.5 times that of the opposite side, or does it refer to the location where the maximum wall thickness is observed?

Reply4: Thank you for your reminder. LVH asymmetry was defined as the maximal LVEDWT > 1.5-fold the wall thickness of the opposing segment. Which has been modified in the manuscript(Methods, Image analysis section).

Changes in the text: We added "maximal" in the sentence "LVH asymmetry was defined as the maximal LVEDWT > 1.5-fold the wall thickness of the opposing segment"(see Page 9, line 4-5).

For Reviewer C

Comment1: From a clinical perspective, differentiating between HCM and HHD is often not particularly challenging. The presence of uncontrolled hypertension, asymmetric hypertrophy patterns, and other clinical signs usually provide clear indicators. Moreover, even if differentiation is challenging with echocardiography alone, heart MRI can reveal characteristic LGE patterns of HCM, obviating the need for advanced techniques like radiomics. Additionally, by managing hypertension and monitoring over time, one can assess LVH regression to further distinguish between HCM and HHD. Given these considerations, it raises the question of why this study specifically emphasized differentiating between HCM and HHD. Were there no considerations given to differentiating other conditions that can cause LVH, such as amyloidosis, uremic cardiomyopathy, Fabry disease, or sarcoidosis? It would be essential for the authors to clarify the clinical rationale behind their chosen focus and explain how this approach provides added value in real-world clinical settings.

Reply1: Thank you for your comments. We acknowledge that distinguishing between HCM and HHD patients with typical presentations is often not particularly challenging. Uncontrolled hypertension, asymmetric hypertrophy patterns, and LGE patterns all contribute to the differentiation of these two diseases. However, not all patients exhibit typical characteristics. For example, in this study, there were as many as 223 HCM patients who were not included due to the presence of concomitant hypertension, accounting for one-third of the total HCM population. Furthermore, more than 40% of HHD patients included in this study showed varying degrees of left ventricular asymmetric hypertrophy, and nearly 30% of HHD patients exhibited enhancement patterns similar to HCM in LGE imaging (mid-wall LGE and RV insertion LGE). A past study has also yielded similar results to our research[1], more than half of the HHD patients with left ventricular hypertrophy exhibited asymmetric hypertrophy (27/50). Additionally, in this subgroup of patients, over 30% showed RV insertion LGE(16/50). Therefore, differentiating these patients is challenging. In fact, models constructed in this study confirmed that models including the Radscore had significantly better discriminatory performance than those without the Radscore, demonstrating the unique value of radiomics in distinguishing between these two diseases.

We have also considered diseases such as amyloidosis, uremic cardiomyopathy, Fabry disease, and sarcoidosis. However, due to limitations in the number of cases, the stability of the radiomics models constructed could not be guaranteed, and therefore, we were unable to conduct research on these diseases at this time. We plan to consider them in the future as part

of a multicenter study with a larger case volume. This has also been added in the limitations of this study.
At the same time, we also acknowledge that we have not provided a detailed explanation of clinical rationale behind our chosen focus and how this approach provides added value in real-world clinical settings.Therefore, we have made revisions to the relevant content in the manuscript(The first paragraph in Introduction section).

Reference:
1.Rodrigues JC, Amadu AM, Dastidar AG, et al. Prevalence and predictors of asymmetric hypertensive heart disease: insights from cardiac and aortic function with cardiovascular magnetic resonance. *Eur Heart J Cardiovasc Imaging*. 2016;17(12):1405-1413. doi:10.1093/ehjci/jev329

Changes in the text: We have modified our text as advised "Although a history of hypertension, symmetrical LVH, indexed LV mass, without right ventricular(RV) insertion late gadolinium enhancement (LGE) and systolic anterior motion of mitral valve (SAM) were found in a previous study to be effective in discriminating HHD from HCM(3). However, not all patients exhibit typical characteristics. RV insertion LGE and asymmetric LVH are not uncommon in HHD patients(4). Hypertension also co-occurs in approximately 40%–60% of adults with HCM(5)(see Page 5, line 5-12).

Comment2: The authors compared the patient characteristics of HCM and HHD within the training set. However, there is no evidence provided to demonstrate that there was no statistical difference between the training and validation sets. To argue for reproducibility, it's essential to ensure that patient group characteristics are consistent in both the training and validation sets.
Reply2: Thank you for your comments. We have added the comparence of patient characteristics in the training and validation sets(Table S2), patient characteristics showed no difference between the training group and the validation group(The first paragraph of the Results section).
Changes in the text: We added"Patient characteristics showed no difference between the training group and the validation group(Table S2)"(see Page 12, line 19-20)

Comment3: The best cutoff value should be determined using AUC in the training group and then applied to the validation group. Subsequently, the sensitivity and specificity can be reported. However, Table 4 shows that AUC was calculated separately for both the training and validation groups, which might not produce the expected outcomes. The authors should report the cut-off, sensitivity, and specificity for the validation group.
Reply3: Thank you for your comments. We have increased the cutoff values for the models construted in this study in the training dataset, and the sensitivity, specificity, and F1 score when applied the cutoff values into the validation set. Additionally, Table 4, which was not the main focus of this paper, has been moved to the supplementary material. In line with your suggestion, a new Table 4 has been added.

Changes in the text: We added "The optimal cut-off value of cine radscore, LGE radscore, combined radscore, clinical model and combined model were -0.40, 0.21, -0.44, 87.81 and 72.02, respectively. The score higher than the cut-off value indicate a higher likelihood of HCM. The sensitivity, specificity, positive predictive value(PPV), negative predictive value(NPV) and F1 score of different models when using the best cutoff values were shown in Table 4"(see Page 14, line 14-19).

Comment4: The discrimination performance should be reported comprehensively, including sensitivity, specificity, and F1 score, not just the AUC.
Reply4: Thank you for your comments. The sensitivity, specificity, positive predictive value(PPV), negative predictive value(NPV) and F1 score of different models when using the best cutoff values were added in Table 4.
Changes in the text: We added "The sensitivity, specificity, positive predictive value(PPV), negative predictive value(NPV) and F1 score of different models when using the best cutoff values were shown in Table 4"(see Page 14, line 17-19).

Comment5: The results of the radiomics feature reproducibility test are missing. It's crucial for the authors to present detailed outcomes of feature selection based on the reproducibility of both intra-observer and inter-observer assessments.
Reply5: Thank you for your comments. The results of the radiomics feature reproducibility test were added in the Results(Reproducibility) section.
Changes in the text: We added"Among the cine images, 1091 features exhibited excellent reproducibility (ICC>0.8 in both intraobserver and interobserver analyses); 269 feature exhibited fair reproducibility(0.8≥the lower ICC value in the intraobserver and interobserver analyses>0.5);49 features exhibited poor reproducibility(intraobserver analyses and interobserver analyses showed at least one ICC≤0.5). Among the LGE images, 1155 features exhibited excellent reproducibility; 121 feature exhibited fair reproducibility; 133 features exhibited poor reproducibility"(see Page 15, line 10-17).

Comment6: The authors should describe the final 20 features in detail and provide plausible explanations regarding how the combination of these features can contribute to discriminating between HCM and HHD.
Reply6: Thank you for your comments. The final 20 features and the formula for calculating Radscore are included in the supplementary materials(Equation S1&S2). These features quantified the differences in the morphology, texture, and voxel distribution of images between the two diseases with or without various filter. The quantitative combination of these differences can help distinguish between these two diseases(In the 7th paragraph of the Discussion section).
Changes in the text: We added" The final selected radiomcis features quantified the differences in the morphology, texture, and voxel distribution of images between the two diseases with or without various filter. Radscores were calculated by the quantitative combination of these differences, which enables the differentiation of the two diseases"(see Page 18, line 12-16).

Comment7: Are there any correlations between radiomics features and MRI parameters? If there is a correlation, it might necessitate a redefinition of the formula used.
Reply7: Thanks for your comment. We conducted Spearman correlation analyses between the final radiomics features and MRl parameters, and found that the correlation coefficients between any MRl parameters and radiomics features were less than 0.5. Therefore, we consider that there is no need to redefinition our formula.
Changes in the text: This part of content has not been added to the article due to length limitations.

Comment8: The comprehensiveness of the reproducibility assessment is questionable. Is a random selection of 30 out of 621 patients representative enough? The authors should cite references to justify their selection method.
Reply8: Thanks for your comment. Reference was cited in the Methods(Reproducibility) section now.
Reference: Ji GW, Zhang YD, Zhang H, et al. Biliary Tract Cancer at CT: A Radiomics-based Model to Predict Lymph Node Metastasis and Survival Outcomes. Radiology 2019;290:90-98.
Changes in the text: We added "(13)" in the sentence "The contours of 30 randomly selected subjects (18 HCM patients and 12 HHD patients) were redrawn after 3 months by observer 1. Observer 2, blinded to the first observer's result, drew the same set of subjects(13)"(see Page 11, line 9-11).

Comment9: Figure 2 lacks clarity. It's crucial to clearly describe the criteria used to select the 20 optimal features.
Reply9: Thanks for your comment. Figure 2 mainly reflects the process of this study. The optimal 20 features have been shown in the supplementary materials.
Changes in the text: Not involve the modification of the content in the article.

Comment10: The manuscript lacks detailed information about radiomics feature extraction. Specifically
- Please categorize the group of 1409 radiomic features. For instance, are they GLCM, GLRLM, wavelets, or other types?
- Were voxels discretized into bins of equal width? Please provide the parameter values used in Pyradiomics.
Reply10: Thanks for your comment. The group of 1409 radiomic features has been added in the first paragraph of Results(Feature extraction and selection) and Methods(Radiomics feature extraction and selection)section. Voxel intensity values were discretized by using a fixed bin width of 25 HU prior to feature extraction. Which has been added in Methods(Radiomics feature extraction and selection) section now.
Changes in the text: We added"voxel intensity values were discretized by using a fixed bin width of 25 HU prior to feature extraction"(see Page 9, line 8-9) and "The radiomics features include semantic, first-, second- and higher-order features. Semantic features include location, shape, size features, etc. First-order features describe the distribution of individual voxels and are median, mean, minimum or

maximum of the intensities on the image. Second-order features describe interrelationships between voxels, which include gray-level size-zone matrix (GLSZM), gray-level co-occurrence matrix (GLCM), gray-level dependence matrix (GLDM), gray-level run-length matrix (GLRLM) and neighborhood gray-tone difference matrix (NGTDM). Higher-order features impose filter grids on the image, which include log, exponential, logarithm, gradient, square, squareroot, wavelet and local binary pattern(LBP)“(see Page 9, line 11-21).

For Reviewer D

Comment1: remove the stard reporting checklist from the introduction, if needed move it in methods section.

Reply1: Thank you for your comment. According to the requirements of the editorial board, this part of the content needs to be placed at the end of the Introduction section.

Changes in the text: Not involve the modification of the content in the article.

Comment2: In methods - study population please rephrase the list of exclusion criteria

Reply2: Thanks for your comment. The list of exclusion criteria has been rephrased.

Changes in the text: We have modified our text as advised **”**Exclusion criteria:1)Patient with other known causes of LVH(e.g., moderate to severe valvular heart disease, amyloidosis, Anderson-Fabry disease, glycogen storage diseases, athletic heart), 2)those with a history of coronary artery disease(>50% stenosis in coronary computed tomography angiography or percutaneous coronary angiography), 3)those coexisting with other heart conditions, such as congenital heart disease or noncompaction of ventricular myocardium, 4)images with significant artifacts or a lack of LGE sequence”(see Page 6-7, line 20-22,1-4).

Comment3: What MLVEDWT is?

Reply3: We are sorry that we did not provide a clear description for mLVEDWT. We have now revised all the relevant content in this article to consistently use "maximal LVEDWT(left ventricular end diastolic wall thickness)" .

Changes in the text: All the “mLVEDWT” in this manuscript has been modified to “maximal EDWT”(see Page 8, line 14, et al).

Comment4: I appreciated limitations included at the end of discussion. Please make this part as a separate paragraph.

Reply4: Thank you for your reminder, which has been integrated in a separate paragraph.

Changes in the text: We have modified our text as advised”This study has several limitations. First, this was a single-center study, and the models built were not validated using external data. Although different scanners were used in this study to extract radiomics features and build discrimination models, the performance of these models showed no significant difference when different equipment was used. Additionally, T1 mapping was not included in our study, which may be useful to differentiate the two diseases, and this will be addressed in future research. Then, genetic test results were not included in this study because they were unavailable for most patients, which will be explored in future studies. Next, due to the high requirements of sample size for radiomics research, this study did not consider the

discrimination of other diseases which may lead to LVH. This will be considered in future multicenter studies. Finally, Due to our center's status as a national hub for cardiovascular diseases in China, patients usually come to our center for treatment only when their condition is severe, the distribution of cases may be uneven"(see Page 19-20, line 10-22,1-2).

Comment5: In introduction please cite: https://doi.org/10.3390/jcm12103481
Reply5: Thank you for your suggestion. Which has been cited in Introduction section(The first sentence, Reference 2).
Changes in the text: We have added reference 2 in the sentence "Hypertrophic cardiomyopathy (HCM) is characterized by left ventricular hypertrophy (LVH) in the absence of another cause of hypertrophy(1-2)"(see Page 5, line 2-3).

Comment6: perform a careful revision of english language
Reply6: Thank you for your comment. This study has been professionally proofread by a specialized company(American Journal Experts, AJE) and a native speaker(Paul Schoenhagen).
Changes in the text: Not involve the modification of the content in the article.