Hypertrophic Cardiomyopathy Diagnosis Based on Cardiovascular Magnetic Resonance Using Deep Learning Techniques

Danial Sharifrazi

Islamic Azad University of Shiraz

Roohallah Alizadehsani (✉ r.alizadehsani@deakin.edu.au)

Deakin University

Navid Hoseini Izadi

Isfahan University of Technology

Mohamad Roshanzamir

Fasa University

Afshin Shoeibi

K. N. Toosi University of Technology

Fahime Khozeimeh

Deakin University

Fariba Alizadeh Sani

Mashhad University of Medical science

Zahra Alizadeh Sani

Omid Hospital

Sadiq Hussain

Dibrugarh University

Chandrashekhar Harlapur

Epworth Hospital

Juan M. Gorriz

Universidad de Granada

Abbas Khosravi

Deakin University

Saeid Nahavandi

Deakin University

Nizal Sarrafzadegan

SPPH, University of British Columbia

Sheikh Mohammed Shariful Islam

Deakin University

Research Article

Keywords: CNN, HCM, Color filtering, Data augmentation

Posted Date: October 26th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1005999/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Hypertrophic cardiomyopathy diagnosis based on cardiovascular magnetic resonance using deep learning techniques

Danial Sharifrazi^a, Roohallah Alizadehsani^{b,*}, Navid Hoseini Izadi^c, Mohamad Roshanzamir^d, Afshin Shoeibi^{e,f}, Fahime Khozeimeh^b, Fariba Alizadeh Sani^g, Zahra Alizadeh Sani^h, Sadiq Hussainⁱ, Chandrashekhar Harlapur^j, Juan M. Gorriz^{k,l}, Abbas Khosravi^b, Saeid Nahavandi^b, Nizal Sarrafzadegan^{m,n}, Sheikh Mohammed Shariful Islam^{o,p,q}

^aDepartment of Computer Engineering, School of Technical and Engineering, Shiraz Branch, Islamic Azad University, Shiraz, Iran

^bInstitute for Intelligent Systems Research and Innovation (IISRI), Deakin University, Geelong, Australia

^cDepartment of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan 84156-83111, Iran

^dDepartment of Computer Engineering, Faculty of Engineering, Fasa University, 74617-81189, Fasa, Iran

^eComputer Engineering Department, Ferdowsi University of Mashhad, Mashhad, Iran

^fFaculty of Electrical and Computer Engineering, Biomedical Data Acquisition Lab, K. N. Toosi University of Technology, Tehran, Iran

^gMashhad University of Medical science, Mashhad, Iran

^hDepartment of Cardiac MRI, Omid Hospital, Tehran, Iran

ⁱSystem Administrator, Dibrugarh University, Assam 786004, India

^jEmergency Department, Epworth Hospital, Geelong, Australia

^kDepartment of Signal Theory, Networking and Communications, Universidad de Granada, Spain.

^lDepartment of Psychiatry, University of Cambridge, UK

^mFaculty of Medicine, SPPH, University of British Columbia, Vancouver, BC, Canada.

ⁿIsfahan Cardiovascular Research Center, Cardiovascular Research Institute, Isfahan University of Medical Sciences, Khorram Ave, Isfahan, Iran

^oInstitute for Physical Activity and Nutrition, School of Exercise and Nutrition Sciences, Deakin University, Geelong, VIC, 3220, Australia

^pCardiovascular Division, The George Institute for Global Health, Newtown, Australia

^qSydney Medical School, University of Sydney, Camperdown, Australia

*Corresponding author: Roohallah Alizadehsani
Email: r.alizadehsani@deakin.edu.au

Abstract

Hypertrophic cardiomyopathy (HCM) can lead to serious cardiac problems. HCM is often diagnosed by an expert using cardiovascular magnetic resonance (CMR) images obtained from patients. In this research, we aimed to develop a deep learning technique to automate HCM diagnosis. CMR images of 37421 healthy and 21846 HCM patients were obtained during two years. Images obtained from female patients form 53% of the collected dataset. The mean and standard deviation of the dataset patients' age are 48.2 and 19.5 years, respectively. Three experts inspected images and determined whether a case has HCM or not. New data augmentation was used to generate new images by employing color filtering on the existing ones. To classify the augmented images, we used a deep convolutional neural network (CNN). To the best of our knowledge, this is the first time CNN is used for HCM diagnosis. We designed our CNN from scratch to reach acceptable diagnosis accuracy. Comparing the designed algorithm output with the experts' opinions, the method could achieve accuracy of 95.23%, recall of 97.90%, and specificity of 93.06% on the original dataset. The same performance

metrics for the designed algorithm on the augmented dataset were 98.53%, 98.70%, and 95.21%, respectively. We have also experimented with different optimizers (e.g. Adadelata and Adagrad) and other data augmentation methods (e.g. height shift and rotation) to further evaluate the proposed method. Using our data augmentation method, accuracy of 98.53% were achieved which is higher than the best accuracy (95.83%) obtained by the other data augmentation methods which have been evaluated. The upper bound on difference between true error rate and empirical error rate of the proposed method has also been provided in order to present better performance analysis. The advantages of employing the proposed method are elimination of contrast agent and its complications, decreased CMR examination time, lower costs for patients and cardiac imaging centers.

Keywords: CNN, HCM, Color filtering, Data augmentation

1 Introduction

In Hypertrophic cardiomyopathy (HCM), the heart muscles abnormally become thick, as illustrated in Figure 1. These thickened muscles make it hard for the heart to pump blood. In some HCM patients, the thickened muscle may lead to shortness of breath. It also may cause chest pain or problems in heart operation which leads to abnormal heart rhythms and even sudden death [1].

Figure 1. In HCM patients, the muscular heart walls i.e. septum are thicker. Left illustrates a normal heart and right shows an HCM heart.

HCM may have one or more symptoms like shortness of breath and chest pain especially during exercise, fainting, heart murmur, and sensation of rapid heartbeats. In most HCM patients, the muscular wall between heart two bottom chambers becomes thicker. Consequently, this

thickened wall may block blood outflow. This problem is called non-obstructive hypertrophic cardiomyopathy if the blocking is not significant. However, the left ventricle may become stiff which makes it hard for the heart to relax. Consequently, it reduces the amount of blood in the ventricle and the amount of blood sent to the body in each heartbeat [2].

Hypertrophic cardiomyopathy patients also have an abnormal arrangement of heart muscle cells known as myofiber disarray. This may trigger arrhythmias in some patients. This disease is usually inherited through families. Having a parent with hypertrophic cardiomyopathy increases the chance of having genetic mutation by 50% [3].

The most common complications of hypertrophic cardiomyopathy are listed as below [2]:

- Atrial fibrillation: this problem i.e. thickened heart muscle alongside the abnormal structure of heart cells may lead to changes in the heart's operation. Consequently, the heartbeats are fast or irregular. Atrial fibrillation increases the risk of blood clots leading to stroke or emboli in other organs or extremities.
- Blocked blood flow: in some patients, the thickened heart muscle may block the blood outflow of the heart. Consequently, it leads to shortness of breath with exertion and other complications such as fainting spells, dizziness, and chest pain [3].
- Mitral valve problems: the valve between the left ventricle and left atrium does not close properly when thickened muscle blocks the blood outflow. Consequently, the blood may leak backward toward the left atrium.
- Weak heart pumping: the thickened heart muscle becomes weak and ineffective, and the ventricle becomes larger, resulting in less forceful heart pumping.
- Heart failure: filling the heart with blood may be disturbed by the stiffness of thickened heart muscle. Consequently, the heart is not able to pump enough blood.
- Sudden cardiac death: this disease may lead to heart-related sudden death in patients of all ages. As many patients do not know they have it, sudden cardiac death may be the first sign of a problem even in seemingly healthy young people or active adults [4].

Based on our knowledge, there is no prevention for HCM. However, it is possible to identify this disease as soon as possible to direct the treatment process and prevent its side effects. For distinguishing the HCM patients, doctors may recommend genetic testing. However, this test may not detect mutation in all HCM patients. In addition, some insurance companies do not cover this test [4].

Echocardiograms can also be used for screening these patients. However, Echocardiograms do not have high accuracy. Therefore, the researchers are trying to find alternative methods for diagnosing HCM based on machine learning and data mining algorithms. To the best of our knowledge, there are few researches in this field [2] which will be reviewed in the next section.

2 Related work

The outstanding representation of convolutional neural network (CNN) has made it one of the most popular machine learning methods among researchers in various fields. Regarding HCM

diagnosis, WY Ko et al. [5] developed a convolutional neural network (CNN) based on ECG signals. The network was trained and validated on 2448 HCM patients and 51153 healthy subjects. To evaluate the trained network, 612 HCM and 12788 healthy subjects were used. The achieved sensitivity and specificity were 93% and 87%, respectively. CNN has also been used for automatic quantification of left ventricle mass in 1073 HCM patients [6]. The reported results demonstrated the power of CNN in automatic segmentation of LV in HCM patients. It was shown that there is no significant difference between automatic and manual segmentation. The same research was done by Q. Tao et al. [7]. In their research, a cine MRI dataset which was obtained from four medical centers was used. Using this dataset, three CNNs with U-NET architecture were trained. Automatically generated results were compared with manually generated ones. Wilcoxon test showed that there is no significant difference between automatic and manually generated results.

Deep learning has also been used to increase the performance of HCM mutation prediction using cardiac cine images [8]. The studied dataset was based on a non-enhanced four-chamber view of cine images and contained information of 198 HCM patients. The classification of the HCM genotypes was realized by a deep learning model. The conducted experiments yielded AUC of 84%, sensitivity of 83.33%, and specificity of 78.26%. Another research employing CNN has been carried out in [9]. The objective was to measure the maximum heart wall thickness of HCM patients. The studied data consisted of 60 adult HCM patients including those carrying HCM gene mutations. Eleven experts and a machine learning algorithm analyzed the maximum heart wall thickness of patients. Compared to the generated results, the precision of the machine learning algorithm was superior.

A. D. Marvao et al. [10] found the prevalence of uncommon variants of HCM genes based on cardiac magnetic resonance imaging of participants. The segmentation of these images was performed using a deep neural network. They reported that due to heart failure and arrhythmia, SARC-P/LP variants were associated with an increased lifetime risk of adverse cardiovascular outcomes. Such variants are also associated with greater risk in HCM patient cohorts. Overall, SARC-P/LP variants have low aggregate penetrance for overt HCM but are associated with an increased risk of adverse cardiovascular outcomes and a sub-clinical cardiomyopathic phenotype.

O. Bernard et al. [11] introduced the “Automatic Cardiac Diagnosis Challenge” dataset, for the purpose of cardiovascular magnetic resonance (CMR) assessment. The dataset contains 150 multi-equipment CMR data. The data were classified by two medical experts. The objective of this work was assessment of state-of-the-art deep learning methods performance in segmenting the myocardium and classifying pathologies.

To the best of our knowledge, none of the existing researches has focused on HCM detection using CMR images and our method is the first in this regard. The novelties of this research are:

- Collecting a new dataset which has been made publicly available to the research community.

- Diagnosing HCM by deep learning algorithm with high accuracy. To the best of our knowledge, it is done for the first time.
- Proposing a new data augmentation method based on color filtering
- Designing a CNN architecture from scratch which is customized to obtain high accuracy on HCM disease instead of using existing models.
- Analyzing the difference between true error rate and the empirical error rate of the proposed method.

Consequently, the following advantages will be achieved:

- Elimination of contrast agent and its complications.
- Shorter examination time of CMR.
- Lower costs for patients and radiology centers.
- Better and faster services by radiology centers.

The rest of this paper is organized as follows. In Section 2, the collected dataset is described. In section 3, the required background for this paper is briefly reviewed. The proposed method is explained in detail in section 4. Implementation details are given in section 5. The experimental results are presented in section 6. Finally, Discussions and conclusions are discussed in section 7.

3 Dataset description

In this study, CMR images of 37421 healthy and 21846 HCM patients were obtained at Omid Hospital in Tehran. They were collected from September 2018 to September 2020 and were labeled by three cardiac imaging experts during the collection time. The mean and standard deviation of their age are 48.2 ± 19.5 years. The percentage of samples collected from female patients is 53%. In parts (a) and (b) of Figure 2, five healthy and five HCM subjects from the collected dataset are illustrated. The collected dataset is publicly available in [12].

Institutional approval was granted by Omid hospital in Tehran so that the patient datasets can be used for research studies on diagnostic and therapeutic purposes. Approval was granted on the grounds of existing datasets. All the patients who have participated in data collection process for this study have completed informed consent form. All the carried out methods complied with relevant guidelines and regulations. Tehran Omid hospital in Iran has given ethical approval in order to use the dataset for research purposes.

Figure 2. Sample CMR images: (a) healthy subjects, (b) HCM patients.

4 Prerequisites

In this section, the required background is briefly reviewed. In section 4.1, HSV color space and its relation to RGB color space are described. CNN is reviewed in section 4.2.

4.1 RGB to HSV conversion

Hue, Saturation, Value (HSV) is an alternative representation for RGB color space. HSV color space represents color more closely to the way human vision perceives color. However, we use HSV instead of RGB for another important reason. Similar to RGB, HSV is made of three channels. Hue channel represents color, S channel is the amount of color, and V channel is the brightness or intensity of the pixel. Therefore, color filtering is done with ease in HSV color space. The desired color interval can be easily set on H channel. However, in RGB space the color filtering involves all three channels which is more complicated compared to HSV space. The conversion from RGB to HSV can be done as follows:

$$\begin{aligned}
 R' &= \frac{R}{255}, G' = \frac{G}{255}, B' = \frac{B}{255}, \\
 C_{min} &= \min(R', G', B'), C_{max} = \max(R', G', B'), \\
 \Delta &= C_{max} - C_{min}, \\
 H &= \begin{cases} 0, & \Delta = 0 \\ 60 \times \left(\frac{G' - B'}{\Delta} \bmod 6 \right), & C_{max} = R' \\ 60 \times \left(\frac{B' - R'}{\Delta} + 2 \right), & C_{max} = G' \\ 60 \times \left(\frac{R' - G'}{\Delta} + 4 \right), & C_{max} = B' \end{cases} \\
 S &= \begin{cases} 0, & C_{max} = 0 \\ \frac{\Delta}{C_{max}}, & C_{max} \neq 0 \end{cases} \\
 V &= C_{max}.
 \end{aligned}$$

4.2 Convolutional Neural Network

Neural network is a machine learning model inspired by how the brain works. A neural network is made of a network of neurons. Neurons are learning units. The parameters of the neurons are determined such that the network input is mapped to the desired output. CNNs are a special type of neural networks which operate on image data [13]. Two key concepts of CNNs are convolutional and pooling layers. The convolutional layers extract key features from the input images via repetitive application of filters on the input images. The output of a convolution layer is a set of activations called a feature map. To make the feature map less sensitive to the location of features in the input image, the feature map is down-sampled using pooling layers. Two common pooling methods are average pooling and max pooling [14]. Combining convolution and pooling, CNNs can extract features from the input images automatically which makes them an ideal choice for various classification and regression applications. In this paper, we use a CNN for binary classification which is trained on Binary Cross Entropy (BCE) loss:

$$loss_{BCE} = \frac{-1}{N} \sum_{i=1}^N [y_i \log(p(x_i)) + (1 - y_i) \log(1 - p(x_i))],$$

where N is the batch size, x_i is the i th input sample, y_i is the i th target label, and $p(x_i)$ is the prediction of CNN about sample x_i .

5 Proposed method

In deep learning, CNN is currently one of the most successful methods with outstanding performance on image classification tasks. However, a huge amount of training data is required to unleash the power of CNNs. In real-world applications, it may be challenging and costly to collect a sufficient number of training samples. It is shown that the CNN classification performance can be improved by increasing the number of training samples via data augmentation methods [13]. Common data augmentation methods include rotating images, shifting images to left or right, zooming in, and changing the brightness of images. In this paper, we propose another data augmentation approach and show experimentally that it can improve CNN performance for image classification. The proposed data augmentation approach is based on the color filtering algorithm. Depending on the characteristics of the problem at hand, filtering may exclude too bright or too dark colors in the image or it may keep a specific color spectrum while discarding the rest of the spectrum from the image.

In each color image represented in RGB color space, the three channels red (R), green (G), blue (B) indicate the intensity of red, green, and blue colors, respectively. When it comes to filtering a specific range of colors from an image, HSV color space yields better results than RGB [15]. To investigate this claim, we have performed our color filtering in both RGB and HSV spaces and compared their performance (Table 1). As will be discussed in section 7.2, converting data color space from RGB to HSV and performing the color filtering in HSV space leads to slightly better performance. That is why we chose to perform our data augmentation (based on color filtering) in HSV color space.

Given the argument above, as the first step of color filtering, the images are converted from RGB color space to HSV (section 4.1). Using the H channel of HSV, the desired color range can be specified for filtering. The portion of image within the specified color spectrum is kept while the rest of the image is removed. As we know, H channel can take on any integer value from 0 to 359. After several experiments, we observed that using interval [58, 198] as the color filtering spectrum yields the best results. The color filtering spectrum is defined around the base point 128 (Figure 3). The lower and upper bounds of the spectrum are $128-70=58$ and $128+70=198$, respectively. Any value outside the defined spectrum ([58, 198]) is discarded during the color filtering.

Figure 3. The color spectrum specified on “H” channel of HSV color space used for filtering

The lower and upper bounds of the filtering spectrum may vary depending on the problem characteristics. For example, the red color may be of interest to us if the detection of red apples is desired. Similarly, the blue color may interest us if our objective is to identify rivers and lakes in satellite imagery.

After applying the color filter on the original images (available ones), they are used as augmented data. Therefore, CNN classifier is trained on original samples as well as the augmented ones. The steps of the proposed method are depicted in Figure 4 which will be explained in more details below:

- As the first step the dataset of images is received.
- The proposed color filter is applied on the dataset samples.
- The dataset is augmented with the filtered images.
- As a preprocessing, the augmented dataset samples are normalized.
- A CNN is trained on the augmented dataset.
- Classification is performed using the trained CNN, and the results are presented.

Figure 4. The high level steps of the proposed method

5.1 Color Filtering Algorithm

The images used in this paper are grayscale, but the proposed color filtering method is also applicable to color images. The objective of color filtering is to discard unwanted details from the images based on their color. The filtering begins by first converting images from RGB to HSV. This is necessary to separate the color and intensity information of the image pixels. The desired color spectrum on Hue is defined using a base color and a threshold around it. For example, in this study, the base color is 128 and the threshold is set to 70. Therefore, the desired color spectrum will be $[128-70, 128+70] = [58, 198]$ as shown in Figure 3. The base color and threshold parameters must be set so that the problem's objective is met. Based on the desired color spectrum, a mask can be formed. By applying the mask on the image, pixels outside the desired color spectrum are discarded while the rest of the pixels are retained. The effect of applying the color filtering algorithm is illustrated in Figure 5.

Figure 5. Illustration of the color filtering algorithm: (a) sample image in RGB space, (b) image converted to HSV, (c) filtering result in HSV color space, (d) filtering result in RGB color space.

6 Implementation details

The color filtering has been carried out using the OpenCV library and Python programming language for image processing. By default, OpenCV loads images with RGB color space which are converted to HSV before applying the color filter.

The classification has been carried out by training a CNN. The training procedure has been implemented using Keras deep learning library with Tensorflow backend. The total number of the dataset (section 3) samples is 59267. The number of healthy subjects is 37421 while the number of HCM patients is 21846. After applying the proposed data augmentation scheme, the number of samples in the augmented dataset increased to 118534 (twice the original dataset). We used 10-Fold cross validation for training since it is the most common method to select model for high dimensional datasets [16]. One fold (11854 samples) was used as test set. The test set samples have been selected randomly from the dataset. The training set was 80% (85344 samples) of the other nine folds (106680 samples) and the remaining 20% (21336 samples) was used as validation set.

The architecture of the CNN is illustrated in Figure 6. As can be seen, the CNN consists of six convolution layers followed by three fully connected layers. Multiple dropout layers have been employed to avoid overfitting.

Figure 6. CNN architecture

7 Experimental results

In this section, experimental results are presented. The upper bound on the difference between true error rate and the empirical error rate is determined in section 7.1 and the proposed data

augmentation approach is evaluated in section 7.2. Finally, the effect of changing the optimizer algorithm on the quality of CNN training is investigated in section 7.3.

7.1 Performance bound analysis

In supervised learning problems, the training and test data are always limited. Therefore, classifiers are always at risk of overfitting to the available data. The true performance bound of the classifiers after training can never be determined exactly. However, it is possible to estimate an upper bound for the performance to get some insight on how well the trained classifier can perform. To this end, consider a training set $\{(x_i, y_i), i = 1, \dots, l\}$ where x_i and y_i are input and desired output for i th sample, respectively. The true error rate of the classifier F with parameter vector $\hat{\alpha}$ for a test sample (x_u, y_u) is defined in terms of probability as [16]:

$$P(\hat{\alpha}) = p(F(\hat{\alpha}, x_u) \neq y_u)$$

where $P(\hat{\alpha})$ is the probability that classifier F misclassifies test sample x_u . The true error rate ($P(\hat{\alpha})$) can never be calculated exactly. However, it is possible to calculate an empirical error rate based on training data (as pointed out in [16]):

$$P_{emp}(\hat{\alpha}) = \frac{1}{l} \sum_{i=1}^l (y_i - F(\hat{\alpha}, x_i)),$$

where $F(\hat{\alpha}, x_i)$ is the prediction of the classifier F for input x_i . As discussed in [16], the difference between true error rate and empirical error rate can be bounded from above as:

$$\gamma_{emp} \leq \sqrt{\frac{1}{2l} \log \left(\frac{N(l, d)}{\eta} \right)}, \quad (1)$$

where η is the significance level and $N(l, d)$ is computed as [16]:

$$N(l, d) = 2 \sum_{k=0}^{d-1} \binom{l-1}{k},$$

where d is the dimension of input x_i . The bound in Equation (1) has been devised for linear classifiers. Our proposed method is based on a CNN which is not a linear classifier. However, as pointed out by Gorriz et al. [16], the last layer of a deep neural network which is a dense layer can be considered as the linear classifier. After all, dense layers are nothing but dot product of weight parameters and outputs from the previous layer. Therefore, in our case, parameter d is equal to the number of neurons in the last layer of CNN in Figure 6 which is 64. Parameter l is also equal to the number of training samples which is 85344 (see section 6). Using Equation (1) and the specified parameters l and d , the upper bound for the performance of the proposed method is plotted in Figure 7. Clearly, as the sample size (number of training samples) increases, the empirical error rate approaches to true error rate and the upper bound on their difference approaches zero. Similarly, reduction of input sample dimension lowers the upper bound.

Considering that in our case $d = 64$ and $l > 80K$, the upper bound on the difference between empirical performance obtained in this paper (see Table 1 and Table 3) and the actual performance is at most 0.05.

Figure 7. Upper bound for difference between true error rate and empirical error rate with 95% confidence level.

7.2 Classification performance on the augmented dataset

To observe the effect of the proposed data augmentation approach, first, the CNN classifier has been trained on the original dataset. Next, the training has been repeated on the augmented dataset. The data augmentation was performed in both RGB and HSV spaces to find out whether using HSV color space is beneficial. The performance of the CNN trained on original data (without augmentation), augmented data using RGB space, and augmented data using HSV space has been shown in Figure 8. The loss and accuracy plots on validation data for the original and the augmented datasets are presented in parts (a) and (b) of Figure 8. As expected, CNN trained on original data has been outperformed by CNN trained on augmented datasets. Moreover, it is observed that performing augmentation in HSV space has led to better performance compared to augmenting data in RGB space.

Figure 8. Performance comparison of CNN trained on augmented data based on color filtering in HSV space (CNN+Aug (HSV)), augmented data based on color filtering in RGB space (CNN+Aug (RGB)), and (CNN) original data: (a) loss plot, (b) accuracy plot.

The effect of training with 10-Fold cross validation with/without augmented data has also been shown by various performance metrics in Table 1. Apparently, regardless of the chosen color space (RGB or HSV), data augmentation has improved the overall classification performance of CNN as observed in the last two rows of Table 1. Comparing the performance of CNN trained on data augmented in RGB and HSV spaces reveals that using HSV space gives minor improvement compared to using RGB space. Based on the argument provided in section 7.1, we can conclude that the variance of the performance measures reported in Table 1 is below 0.05.

Table 1. Performance metrics of CNN classifier for training with and without data augmentation.

	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-score (%)	AUC (%)
Performance on original dataset	95.23	95.50	97.90	93.06	96.90	95.05
Performance on augmented dataset with RGB-based color filtering	96.62	97.19	98.03	94.01	96.81	95.69
Performance on augmented dataset with HSV-based color filtering	98.56	97.00	98.70	95.21	98.00	96.67

7.3 Comparison of applying different optimizers

In our experiments, for CNN training, we tried different optimizers such as Adadelata [17], Adagrad [18], and Adam [19]. The performance of these methods has been summarized in Table 2. As can be seen, Adadelata has performed better than Adagrad. The superiority of Adadelata stems from the fact that it addresses the aggressive learning rate reduction of Adagrad [17]. Among the methods listed in Table 2, Adam has the best performance since it combines the advantages of two popular optimization methods namely Adagrad and RMSProp[19]. The loss

and accuracy plots (on validation data) during training CNN with the three optimizers (Figure 9) agree with the performance metrics presented in Table 2. Adam has the least loss and the best accuracy. The second best is Adadelta and Adagrad takes third place.

Table 2. Performance comparison between different optimizers.

Optimizer	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-score (%)	AUC (%)
Adadelta	95.71	95.30	98.10	92.63	96.60	94.89
Adagrad	94.98	94.5	97.50	91.79	96.10	94.01
Adam	98.53	97.00	98.70	95.21	98.00	96.67

Figure 9. Comparison of different optimizers for CNN training: (a) loss plot, (b) accuracy plot.

7.4 Comparison of different data augmentation methods

To compare the proposed data augmentation method with existing ones, we have augmented the dataset with height shift [20], rotation [21], and width shift [22]. The results of training the CNN on the data augmented with these methods are presented in Table 3. The best performance has been achieved by the proposed augmentation method. Second best is rotation augmentation. Height and width shifts take the third and fourth place, respectively. Based on the argument provided in section 7.1, we can conclude that the variance of the performance measures reported in Table 3 is below 0.05. The loss and accuracy plots of CNN trained on dataset augmented using the methods in Table 3 are illustrated in Figure 10. Clearly, the proposed augmentation has led to better accuracy and lower loss compared to the rest of the augmentation methods.

Table 3. Performance comparison between different data augmentation methods.

Augmentation method	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-score (%)	AUC (%)
Proposed	98.53	97.00	98.70	95.21	98.00	96.67
Height shift	94.45	94.70	96.80	93.65	95.70	93.63
Rotation	95.83	95.20	98.20	94.20	97.00	94.95
Width shift	92.00	91.70	96.00	90.06	93.70	90.56

Figure 10. Comparison of different data augmentation methods: (a) loss plot, (b) accuracy plot.

7.5 Grad-CAM results

Grad-CAM is a generalization of the Class Activation Mapping. It does not need retraining and can be applied broadly to any CNN-based architectures [23]. Some samples of CMR images and their corresponding Grad-CAM results are shown in Figure 11. Three CMR images for normal cases and the result of applying Grad-CAM on them are shown in parts (a) and (b) of Figure 11. Similarly, three CMR images for HCM cases and their Grad-CAM results are presented in parts (c) and (d) of Figure 11.

(a) Normal cases

(b) Grad-CAM for normal cases

Figure 11. (a) examples of normal CMR images. (b) result of Grad-CAM for normal cases. (c) examples of CMR images for HCM cases. (d) result of Grad-CAM for HCM cases.

8 Discussions

HCM patients may have no explicit symptoms, which makes the diagnosis process challenging for experts. In this research, we proposed an automatic method for HCM diagnosis using a CNN trained on CMR images and achieved high accuracy. Widespread use of this method may help experts and patients for better decision-making to treat this disease while reducing the time and cost for patients and radiology centers. However, more studies are required for this method to be used in clinical practice.

Diagnosing HCM by human experts is prone to mistakes. The ensemble of human experts and the proposed method may lead to more accurate decision-making about HCM patients and prevent severe complications of misdiagnosis. The proposed system performance could be improved provided that more training is done on newly collected images. Despite having high accuracy, the proposed system must be validated by medical practitioners.

Machine learning algorithms have shown excellent performance in the health care domain. However, to put these methods to some good use, gaining the medical experts' trust is a must. Therefore, our final goal is to use the proposed method as a tool in routine clinical care.

In the medical field, disagreement between experts' opinions is common. Under such circumstances, the experts have no choice but to conduct more tests on patients to make more accurate decisions. More tests mean higher treatment costs and time. However, it is possible to exploit the potential of machine learning algorithms and use them as a source of decision-

making. Combining the opinion of human experts with the decision of the algorithms may lead to better diagnosis with fewer tests and faster.

In this research, a new CNN architecture was designed from scratch to obtain high accuracy on HCM disease diagnosis instead of using existing models. Our machine learning tool not only shortens CMR examination time but also lowers the costs for patients and radiology centers. Furthermore, the performance of our method can be improved using new training data collected over time.

To the best of our knowledge, no existing work has used CNN for HCM diagnosis using CMR images. The advantage of our method is its automation that can assist the medical experts with their decisions. It eliminates the contrast agent and its complications, shortens the examination time of CMR, and lowers the costs for patients and cardiac imaging centers. This leads to better and faster services by cardiac imaging centers.

Our proposed system has also utilized a new data augmentation approach and showed experimentally that it could improve CNN performance for image classification. Our data augmentation approach is based on the color filtering algorithm. After applying the color filter on the original images, they were used as augmented data. Therefore, the CNN classifier was trained on original samples as well as the augmented ones. For analyzing the performance of our method, we experimented with different optimizers and data augmentation methods. The results showed that our designed algorithm outperforms all evaluated data augmentation methods. During our experiments, we observed that performing the color filtering in HSV space yields better results compared to performing the filtering in RGB space which is why we use HSV space for our data augmentation based on color filtering.

Unfortunately, our research has some limitations. It is not fully automated, as an expert still needs to have a quality control role. All data were collected from a limited region (one hospital). Having more data collected from various health care centers may improve the reliability of our algorithm. Additionally, this method has not been tested with poor-quality images. These images are generated, for example, when a patient cannot hold their breath or those with implantable devices or atrial fibrillation.

In conclusion, we showed how a machine learning tool could be used for HCM diagnosis more precisely. Delivering high accuracy is the first requirement of any machine learning method for HCM diagnosis. As the next step, the method must be validated in practice on new cases. A tool with more precise prediction compared to experts is the ultimate goal of learning methods. Widespread usage of such a tool could help to improve HCM diagnosis. Consequently, sudden cardiac death risk will be decreased, leading to higher life expectancy in HCM patients.

Data sharing

The dataset used in this work has been released for public use in [12].

Author contributions

RA, SH, AS, FAS, NHI, and FK had continuation to prepare first draft. ZAS, NS, AK, SMSI, JMG, and SN contributed to edit final draft. RA, SH, NHI, CH, DS, AS, and MR contributed to all analysis of the data and produced the outcomes accordingly. DS, AS, CH, RA, FAS, FK and MR found the papers and then extracted data. ZAS, JMG, NS, SMSI, AK, and SN provided overall guidance and managed the project.

Competing interests

The authors declare no competing interests.

Acknowledgement

There is not acknowledgement to declare.

References

- [1] B. J. Maron and M. S. Maron, "Hypertrophic cardiomyopathy," *The Lancet*, vol. 381, no. 9862, pp. 242-255, 2013/01/19/ 2013, doi: [https://doi.org/10.1016/S0140-6736\(12\)60397-3](https://doi.org/10.1016/S0140-6736(12)60397-3).
- [2] "<https://www.mayoclinic.org/diseases-conditions/hypertrophic-cardiomyopathy/symptoms-causes/syc-20350198>," 2021.
- [3] B. J. Maron, R. O. Bonow, R. O. Cannon III, M. B. Leon, and S. E. Epstein, "Hypertrophic cardiomyopathy," *New England Journal of Medicine*, vol. 316, no. 14, pp. 844-852, 1987.
- [4] B. J. Maron, "Hypertrophic CardiomyopathyA Systematic Review," *JAMA*, vol. 287, no. 10, pp. 1308-1320, 2002, doi: 10.1001/jama.287.10.1308.
- [5] W.-Y. Ko *et al.*, "Detection of Hypertrophic Cardiomyopathy Using a Convolutional Neural Network-Enabled Electrocardiogram," *Journal of the American College of Cardiology*, vol. 75, no. 7, pp. 722-733, 2020/02/25 2020, doi: 10.1016/j.jacc.2019.12.030.
- [6] S. Fahmy Ahmed *et al.*, "Automated Cardiac MR Scar Quantification in Hypertrophic Cardiomyopathy Using Deep Convolutional Neural Networks," *JACC: Cardiovascular Imaging*, vol. 11, no. 12, pp. 1917-1918, 2018/12/01 2018, doi: 10.1016/j.jcmg.2018.04.030.
- [7] Q. Tao *et al.*, "Deep Learning-based Method for Fully Automatic Quantification of Left Ventricle Function from Cine MR Images: A Multivendor, Multicenter Study," *Radiology*, vol. 290, no. 1, pp. 81-88, 2019, doi: 10.1148/radiol.2018180513.
- [8] H. Zhou *et al.*, "Deep learning algorithm to improve hypertrophic cardiomyopathy mutation prediction using cardiac cine images," *European Radiology*, 2020/11/25 2020, doi: 10.1007/s00330-020-07454-9.
- [9] J. B. Augusto *et al.*, "Diagnosis and risk stratification in hypertrophic cardiomyopathy using machine learning wall thickness measurement: a comparison with human test-retest

- performance," *The Lancet Digital Health*, vol. 3, no. 1, pp. e20-e28, 2021/01/01/ 2021, doi: [https://doi.org/10.1016/S2589-7500\(20\)30267-3](https://doi.org/10.1016/S2589-7500(20)30267-3).
- [10] A. de Marvao *et al.*, "Outcomes and phenotypic expression of rare variants in hypertrophic cardiomyopathy genes amongst UK Biobank participants," *medRxiv*, 2021.
 - [11] O. Bernard *et al.*, "Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved?," *IEEE Transactions on Medical Imaging*, vol. 37, no. 11, pp. 2514-2525, 2018, doi: 10.1109/TMI.2018.2837502.
 - [12] "<https://www.kaggle.com/danialsharifrazi/hypertrophic-cardiomyopathy-dataset>."
 - [13] D. Sadeghi *et al.*, "An Overview on Artificial Intelligence Techniques for Diagnosis of Schizophrenia Based on Magnetic Resonance Imaging Modalities: Methods, Challenges, and Future Works," *arXiv preprint arXiv:2103.03081*, 2021.
 - [14] M. Khodatars *et al.*, "Deep Learning for Neuroimaging-based Diagnosis and Rehabilitation of Autism Spectrum Disorder: A Review," *arXiv preprint arXiv:2007.01285*, 2020.
 - [15] "https://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_imgproc/py_colorspaces/py_colorspaces.html#object-tracking," 2021.
 - [16] J. M. Górriz, J. Ramirez, and J. Suckling, "On the computation of distribution-free performance bounds: Application to small sample sizes in neuroimaging," *Pattern Recognition*, vol. 93, pp. 1-13, 2019/09/01/ 2019, doi: <https://doi.org/10.1016/j.patcog.2019.03.032>.
 - [17] M. D. Zeiler, "Adadelta: an adaptive learning rate method," *arXiv preprint arXiv:1212.5701*, 2012.
 - [18] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *Journal of machine learning research*, vol. 12, no. 7, 2011.
 - [19] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
 - [20] L. F. Sánchez-Peralta, A. Picón, F. M. Sánchez-Margallo, and J. B. Pagador, "Unravelling the effect of data augmentation transformations in polyp segmentation," *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, no. 12, pp. 1975-1988, 2020/12/01 2020, doi: 10.1007/s11548-020-02262-4.
 - [21] J. Shijie, W. Ping, J. Peiyi, and H. Siping, "Research on data augmentation for image classification based on convolution neural networks," in *2017 Chinese Automation Congress (CAC)*, 20-22 Oct. 2017 2017, pp. 4165-4170, doi: 10.1109/CAC.2017.8243510.
 - [22] C. Shorten and T. Khoshgoftaar, "A survey on image data augmentation for deep learning. *J. Big Data* 6 (1), 1–48 (2019)," ed.
 - [23] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618-626.