Detection of Myocardial Infarction Using Computer Vision and Deep Learning Techniques

Frank Cally A. Tabuco*
fatabuco@up.edu.ph
University of the Philippines Diliman
Quezon City, Metro Manila, Philippines

Abstract

Myocardial infarction is a leading cause of mortality worldwide. Traditional detection of this disease relies on the use of echocardiographs which are then analyzed by physicians based on motion of the Left Ventricular (LV) wall in apical 4-chamber view. However, a major challenge in its detection is the quality of extracted echos which, in the worst case, exhibits high degree of noise, blurriness, and incomplete regions. In this paper, we tackle the problem of detecting myocardial infarction in low-quality echocardiograms using the recently released Hamad Medical Corporation and Qatar University benchmark dataset. The contributions of this paper are as follows: 1) Use classical image enhancement techniques to improve segmentation results of the LV wall with low variation; 2) Apply handcrafted feature descriptors using Harris corner detector and Histogram of Oriented Gradients; and 3) Use of data augmentations to improve generalizability of deep learning models on the limited number of echo frames. Our work outperformed the baseline results for the segmentation task yielding scores of 99.76%, 94.68%, 94.64%, and 99.54% for specificity, precision, F1, and accuracy, respectively. On the classification task, ResNest50 yielded the best specificity score with 96.64%.

CCS Concepts: • **Computing methodologies** → **Support vector machines**; **Image segmentation**; **Matching**.

Keywords: computer vision, harris detector, medical image segmentation, myocardial infarction

1 Introduction

Globally, the leading cause of mortality is a group of diseases affecting the cardiovascular system. Cardiovascular diseases (CVDs) include several heart and blood vessel complications such as cerebrovascular disease, rheumatic heart disease, coronary heart disease [20], or invasive diseases such as cardiac amyloidosis [22]. According to the World Health Organization (WHO), this group of diseases accounted for 31% of worldwide deaths in 2019 and is now becoming common even in developing countries [30]. However, most deaths related to CVD is caused by myocardial infarction amounting to 80% of reported deaths [20]. Myocardial infarction (MI), or most commonly known as heart attack, is a condition in which supply and demand of blood through the heart is blocked causing a wide range of symptoms including nausea, chest tightness, and jaw discomfort [27]. These blockages are caused by the formation of plaques which are often residual build-ups of fat and other substances [30]. In previous years, initial evaluation of CVDs tends to rely on clinical suspicion through the onset of chest discomfort and other symptoms [2, 18]. Early detection of myocardial infarction and other CVDs utilize diagnostic tools such as electrocardiograph and echocardiograph [19, 22, 27] which provide robust pathological information such as cardiac voltage output, heart wall thickness, diastolic dysfunctions, and extracellular spaces [22, 27].

Early detection of diseases is crucial to reduce patient mortality rates even for diseases which are terminal in later stages [3, 31]. Advances in medical technologies, which paved way for noninvasive image extraction techniques, help in timely detection. Echocardiographs, magnetic resonance imaging (MRI) scans, and computed tomography (CT) scans are some of the modalities now being utilized by radiologists and physicians to analyze body abnormalities [14, 22, 26, 27]. With these modalities, medical image datasets are now becoming more common. Coincidentally, advances in computer vision algorithms, and deep learning architectures have created a branch of research which focuses on medical image analysis. These new datasets combined with state-of-the-art machine learning algorithms have helped in medical image tasks such as segmentation, classification, and prediction [6, 29, 32].

Two sub-problems are usually included in detection tasks: segmentation and classification. In this paper, we utilize the Hamad Medical Corporation and Qatar University (HMC-QU) benchmark dataset to tackle the detection problem for medical images [9, 13]. This dataset consists of echocardiogram videos in low-resolution, ground-truth segmentation masks, and annotated labels for LV wall segments with MI. For the segmentation problem, we use a variant of U-Net architecture [34] along with different image enhancement techniques to achieve the best performance against ground truth masks with low variance. For the classification problem, we use three techniques employing image stitching

*Main corresponding author

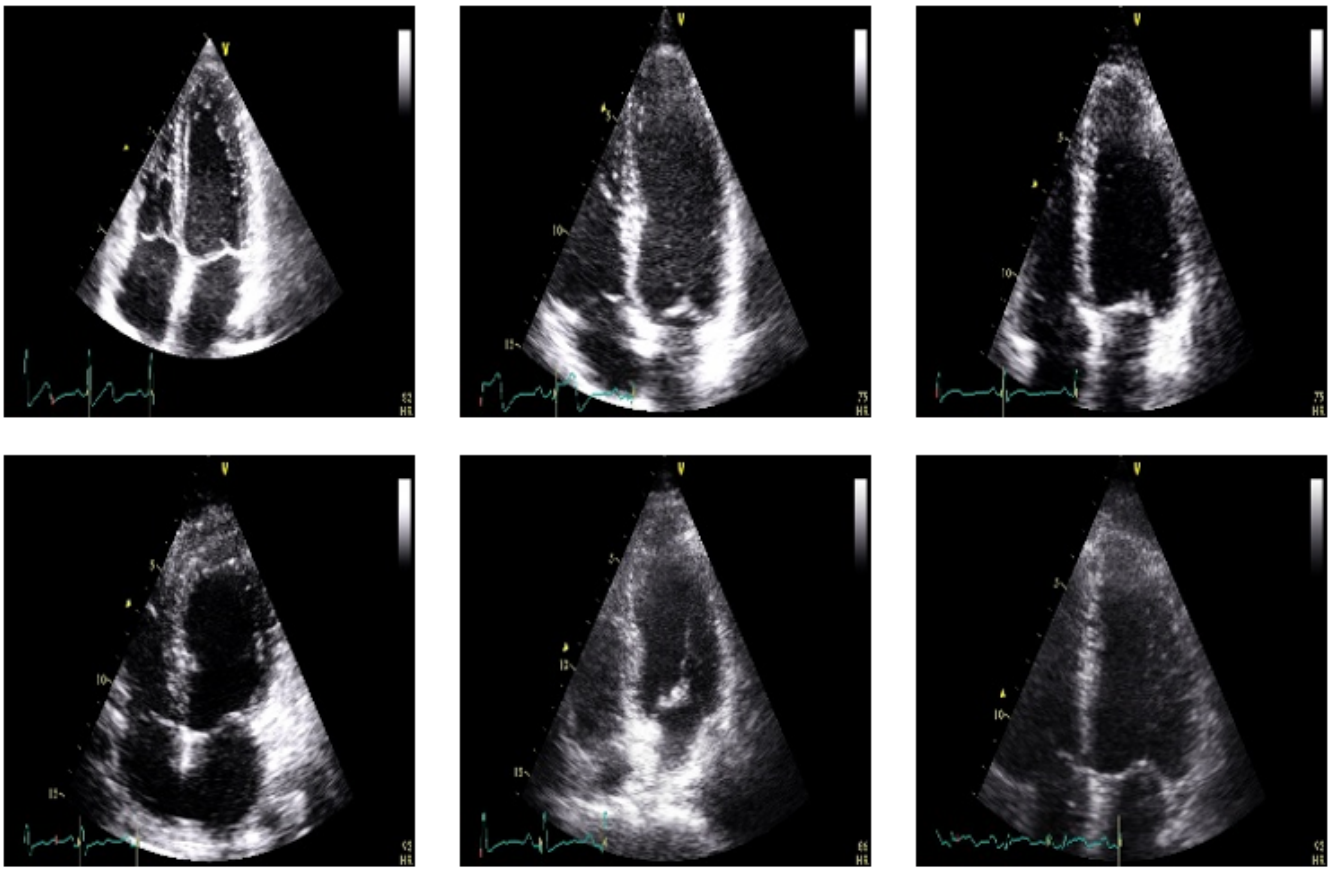

Figure 1. Sample echo frames. Each frame has different qualities with images on center top and bottom right exhibiting high amount of noise.

using Harris detectors [11], Histogram of Oriented Gradients (HOG) [8], and convolutional neural network.

The rest of the paper is structured as follows: section 2 will discuss some related literature for the detection task in medical image analysis. The methods employed for segmentation and classification are discussed in detail in section 3. Section 4 gives an overview of the experiment design starting from extracting echo frames from videos, splitting the dataset, and employing image enhancement techniques to training segmentation networks, and classifiers. Included in this section is a discussion of the parameters set for the experiments. In section 5, we present the results of the experiments. Finally, the last sections discusses improvements for future works.

2 Related Works

Medical image analysis is a branch of research in machine learning with a multitude of real-world applications. The aim of this research is to help physicians in improving patient survival through accurate and timely diagnosis of diseases. However, one major disadvantage of medical image analysis is the poor quality of extracted images. Medical images are often noisy and blurry making it difficult even for skilled radiologists to identify regions with abnormalities, and could prevent them from giving proper diagnosis [14, 15]. Therefore, image enhancements are often used to provide better image qualities which reflect better features that are useful to improve generalization of machine learning models [15]. In the work by Agravat and Raval [1], they proposed pre-processing of images prior to model training to significantly improve results. In their work, they used noise removal, intensity normalization, and histogram matching for segmenting brain images with glioma tumors.

Image enhancement is often paried with segmentation tasks but not always necessary. For image segmentation of LV walls, Degerli et. al [9] proposed an Encoder-Decoder convolutional neural network (E-D CNN) to extract the LV wall in apical 4-chamber (A4C) view in echocardiograms. Their proposed network, which is similar to a U-Net architecture, achieved accuracy and sensitivity scores of 99.42% and 95.72%, respectively. They were able to achieve these scores even without employing image enhancement techniques for the noisy 2D frames.

In [10], they used Harris corner detector in determining key points in brain medical images. The extracted corners were then used to create importance-values necessary for diagnosis of brain disorders. For MI detection, [9] classified the segments based on infarcted and healthy regions. They used handcrafted features to determine LV wall motion such as displacement of endocardial boundary points, displacement of segment centers, and intersection of segments across 2D echo frames. These features were then used as inputs for several classical machine learning classifiers with Support Vector Machines (SVM) achieving the best scores of 80.24% and 85.97% for accuracy and sensitivity, respectively.

3 Methodology

3.1 Dataset

The HMC-QU benchmark was released to the public by [9, 13] in early 2021. The dataset includes a total of 160 anonymous 2D A4C echos obtained from 2018 and 2019. Of the total echo videos, only 109 have ground truth segmentation masks, therefore only this subset will be used for LV wall segmentation and myocardial infarction detection. For the LV wall segmentation task, frames within the end-diastole and end-systole were extracted which varies between 11 to 30 frames per echo. Only these extracted frames have ground truth masks amounting to a total of 2,345 frames for the 109 echos. Sample echo frames from the dataset are shown on Figure 1. In [9], the segmented LV wall was further split into 7 regions as seen in Figure 2. However, for the classification task, only 6 segments have ground truth labels since segment 4 shows no movement activity across all frames in an echo and thereby excluded in the analysis. The classification task is framed as a multilabel problem where an echo can consist of multiple combinations of infarcted and healthy segments. In this work, we use the results from [9] as baselines for the segmentation and classification tasks.

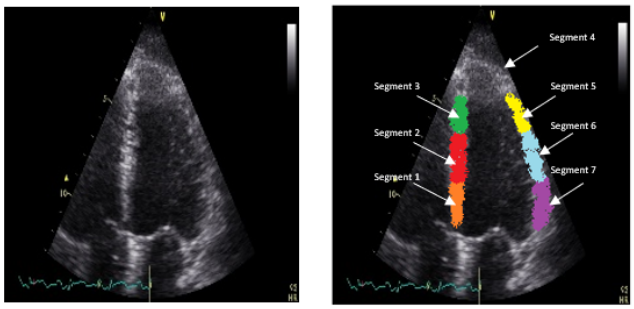

Figure 2. Segment regions of the LV wall.

3.2 Image Enhancement

Before feeding the extracted frames into the segmentation network, we perform some image enhancement techniques to stabilize the noise in each echo. In figure 1, the level of noise is not the same for all echos. As such, a suitable image enhancement which will improve segmentation performance with low variance across all frames is essential to determine the most appropriate technique for low-quality, noisy echocardiograms. We use four image enhancement techniques to equalize and remove noise across frames. First, we use Gaussian blurring with a kernel size of 3x3. Second, we apply Otsu thresholding with a binary-threshold value of 0 and a maximum pixel value of 255. For the third technique, we use gray-level slicing. We can see in figure 2 that the LV wall pixels have stronger intensities than surrounding regions of the heart. As such, we hypothesize that by removing lower intensity pixels the segmentation model's performance would improve. We first compute for the mean pixel value of a frame and further divide it by a value of 1.5 to account for frames with low overall pixel intensity. The gray-sliced image is composed of pixel values which are greater than the minimum set value. Finally, we use a combination of two techniques where an image is first enhanced using Gaussian blurring and then perform gray-level slicing.

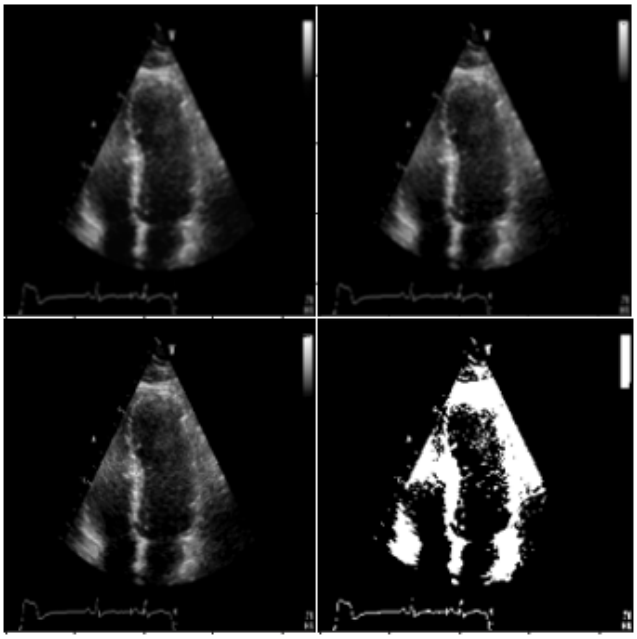

Figure 3. Image enhanced images. From top-left, clockwise: Gaussian blur, gray-level slicing with Gaussian blur, Otsu thresholding, and gray-level slicing.

3.3 Segmentation

3.3.1 U-Net++. There are several architectures and models used for image segmentation which includes the use of convolutional neural networks, fully convolutional networks, and autoencoder architecture. Most popular of among these is the U-Net architecture created by Ronnerberger, Fischer, and Brox [23]. Their network relies on a contracting path,

which resembles a normal CNN architecture, and a expansive path, which up-samples features using a 2x2 convolution. Features from the contracting path are also cropped and concatenated to features in the expansive path. In this paper, we use a variant of U-Net named, U-Net++. Zhou et. al [34] created this variant which outperformed the base architecture in terms of Intersection over Union (IoU) score by as much as 6%. The architecture uses dense skip connections while adding an up-sampling convolution layer for each skip pathways to improve the flow of gradients. Their architecture is shown in Figure 4 which shows their novel approach of using convolution layers in skip pathways.

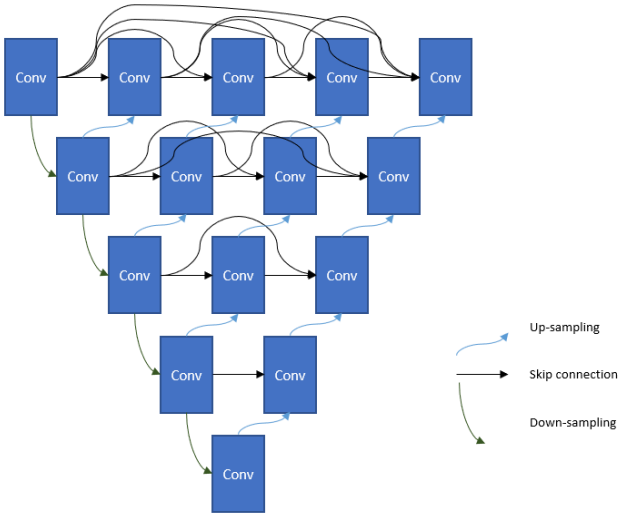

Figure 4. U-Net++ architecture.

3.4 Classification

3.4.1 Harris Corner Detection. After computing for the LV wall segment predictions, we apply the predicted mask on to the original echo frame to produce segmented frames. Each segmented frame is then used to extract reference points for each of the 6 segments of the LV wall. This follows a similar approach in [9] where they computed for the maximum displacement of certain points for each frame in an echo. However, instead of using an average of endocardial boundary points, we only extract three points per segment corresponding to the first, middle, and last pixels of a segment. The extracted points are then compared against each frame to determine their new locations by using Harris corner detection [11] and homography matching. Once the matched points have been determined, we comput for the L2 norm between these points and store the maximum distances for each segment points. The L2 norm between points is computed using the formula:

$$||a||_2 = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (1)$$

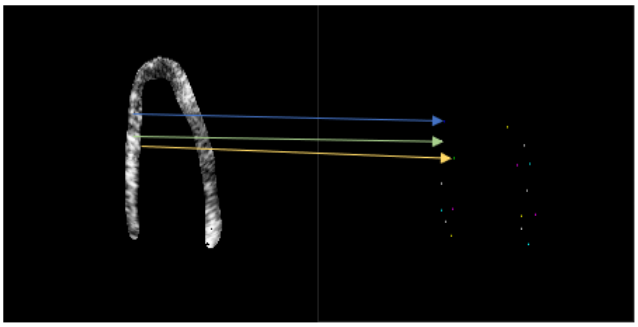

Figure 5. Left: sample segmented frames based on predicted masks of LV wall. Right: extracted points per segment

The images in Figure 5 and 6 depict a sample implementation of feature matching using three points per segment. A total of 18 maximum displacement points are computed between each frames based on best matches.

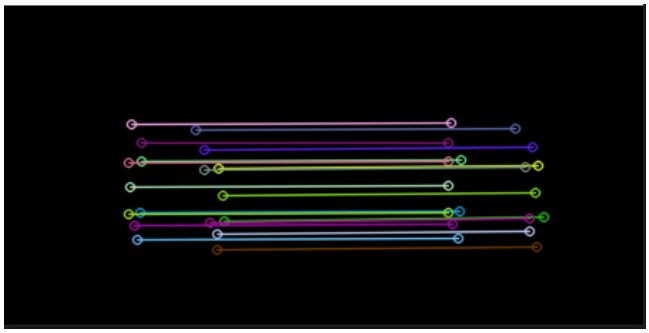

Figure 6. Feature matching with homography.

3.4.2 Histogram of Oriented Gradients. Histogram of Oriented Gradients (HOG) is a feature descriptor developed for human and object detection [8]. This feature engineering technique extracts gradients and orientations through calculation of localized portions of an image. HOG then generates a Histogram for each of the pixel features. Aside from human and object detection, HOG has also been used for motion detection in 3D magnetic resonance image (MRI) scans [24]. Their approach extends the implementation of HOG from 2D images to 3D images by first extracting image gradients then placing these gradients into uniformly-spaced bins.

In this work, we use HOG to detect motion in echocardiograph frames. The approach builds on the extracted segmented frames from the segmentation network then computes HOG descriptors for each frame. Since frames for each echo have varying lengths, we extract five uniformly spaced

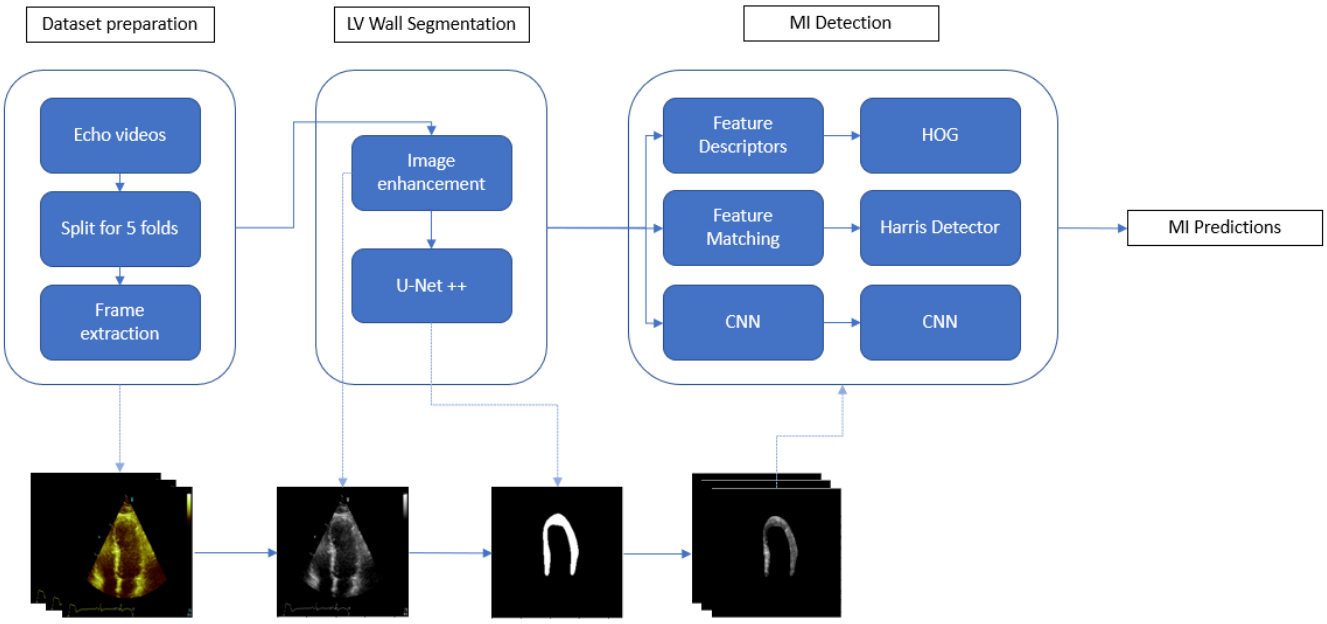

Figure 7. Overview of experiment methodologies for segmentation and classification.

segmented frames. Then, we compute for each HOG descriptors and concatenate the results to form a 9,408 echo feature descriptor vector.

3.4.3 Convolutional Neural Networks. Most medical image analysis problems comes with three common challenges. First is the lack of available data especially for rare types of diseases. It takes years to produces datasets with sufficient amount of samples due to delays cause by extraction of images, pre-processsing datasets, and authorization of datasets from stakeholders [9]. Aside from data availability, these datasets are also imbalanced with more samples showing healthy regions than those with abnormalities. Finally, medical images are often noisy making it difficult even for skilled physicians to identify regions with abnormalities [14, 15]. With only 2,349 frames, more classes with healthy segments (444 healthy compared to 210 infarcted segments), and low-quality images, the HMC-QU dataset has all these challenges. These disadvantages significantly affect the performance of convolutional neural network (CNN) models which are heavily reliant on high-quality, big data. To circumvent this, we employ data augmentations during CNN training to improve generalizability and performance on relevant metrics. Image augmentation changes the appearance of images in order to highlight or remove certain image features [1, 16, 17, 32]. We use the albumentations library [5] to implement the following augmentations:

- Random rotation from -90° to 90° [25]
- Gaussian noise
- One of the following pixel augmentations: 1) contrast limited adaptive histogram equalization, 2) random brightness, or 3) random gamma
- One of the following blurring and sharpening: 1) sharpen, 2) blur, or 3) motion blur

For the CNN network, we chose a pre-trained ResNest50 as the backbone network of our CNN. Split-Attention Networks or ResNest uses split-attention blocks where different feature maps of an image are correlated together while still preserving each feature's independent representations [33]. We attach a fully connected layer at the final CNN layer which outputs a binary multi-label class, one value for each of the six segments.

4 Experiment Design

An overview of the experiment design is shown on Figure 7. For our work, we split the dataset consisting of 109 echo videos into 80% training set, 10% validation set, and 10% test set using 5-fold stratified cross validation. After splitting, we extract the frames within the annotated end-diastole and end-systole duration. Image enhancement techniques discussed in section 3.2 are then applied to the extracted frames using OpenCV [4]. For the segmentation task, the U-Net++ network uses pre-trained weights on imagenet with a Sigmoid activation function, Adam optimizer, a learning rate of 0.001, and a Dice Loss function. We train the model for 25 epochs and store the best model with the lowest validation dice loss.

The best model trained by the U-Net network is then used to predict masks of the LV walls. Each mask is then added

Table 1. Segmentation performance results on 5-fold cross validation.

Model	Sensitivity	Specificity	Precision	F1	Accuracy
[9]	**95.72 ± 1.69**	99.58 ± 0.23	91.11 ± 4.25	93.29 ± 1.77	99.42 ± 0.18
U-Net++	94.84 ± 1.02	99.74 ± 0.07	94.30 ± 1.61	94.49 ± 0.41	99.53 ± 0.04
Blur	94.84 ± 0.52	99.75 ± 0.04	94.46 ± 1.00	94.58 ± 0.31	99.52 ± 0.03
Grayslice & blur	94.54 ± 0.78	99.75 ± 0.04	94.49 ± 0.97	94.45 ± 0.16	99.53 ± 0.01
Grayslice	94.75 ± 0.95	**99.76 ± 0.05**	**94.68 ± 1.24**	**94.64 ± 0.32**	**99.54 ± 0.03**
Otsu	92.28 ± 1.65	99.72 ± 0.09	93.62 ± 2.03	92.81 ± 0.30	99.39 ± 0.03

Table 2. Detection performance results on 5-fold cross validation.

Model	Sensitivity	Specificity	Precision	F1	Accuracy
SVM[9]	**85.97**	70.10	85.52	85.29	**80.24**
RF[9]	80.26	71.81	**85.99**	**82.57**	77.47
Harris-SVM	47.16	87.39	63.65	54,12	74.45
Harris-RF	42.35	83.56	54.89	47.65	70.32
HOG-SVM	38.78	84.80	54.69	45.11	69.99
HOG-RF	25.24	88.02	50.36	33.52	67.86
CNN	23.46	**96.64**	70.52	34.50	78.63
CNN-LSTM	35.02	89.54	52.73	41.45	75.45

Table 3. Harris results after applying balancing techniques.

Model	Sensitivity	Specificity	Precision	F1	Accuracy
Harris-SVM	73.85	72.36	55.76	63.30	72.76
Harris-RF	51.47	79.90	55.10	53.10	70.76

to the original echo frame to produce segmented frames as seen on Figure 5. For the MI classification task, we use three techniques employing classical computer vision and deep learning. First, the Harris corner detector is computed up to subpixel accuracy with a thresholding level set at 0.0001 and other parameters set to default values in OpenCV [4]. For HOG descriptors, we used the built-in function in scikit-image [28] with orientations set to 8, pixels per cell of 16x16, and cells per block of 1x1. Both of these extracted features are used as inputs for Random Forest and SVM classifiers using sklearn libraries [21]. Our work sets 10 estimators for Random Forest, and used RBF kernel with a cost value of 10 for SVM. Lastly, we use a ResNest model combined with data augmentations during training. This network uses an Adam optimizer with a learning rate of 0.001, a BCEWithLogitsLoss function, and a Sigmoid activation function. The model is trained for 25 epochs while storing the model with the highest validation accuracy score.

5 Results & Discussion

We compare our methods with the base results in [9] for both segmentation and classification tasks. Table 1 summarizes the average scores and standard deviation results for the segmentation task. It shows that the encoder-decoder architecture created by Degerli et. al [9] yielded the best overall score for sensitivity with an average of 95.72%. Remarkably, they were able to achieve this result even without employing any image pre-processing. However, much like their result in sensitivity, performance on other metrics shows high variation especially for Precision and F1 score.

Our method, which adds simple image enhancements before training the segmentation network, outperformed the baseline results for specificity, precision, F1, and accuracy by as much as 3.57%. Deviations from the average scores are also much lower with Precision exhibiting the highest change from 4.25% to 1.24%. The results imply that simple pixel augmentations improves generalizability of the network to unknown data. Of the enhancement techniques we used, gray-level slicing yielded the best average scores. Gray-level slicing removes pixels with intensities lower than the minimum set value. This minimum value is set to equal the mean pixel intensity for a given frame multiplied by a hyperparameter set to 0.6667. This dynamic approach helps in stabilizing noise for low-quality echocardiography especially for images where the LV wall is seemingly blending with other regions of the heart (see Figure 1, bottom-right image).

Despite high results for the segmentation task, our methods for the classification task pale in comparison to handcrafted features developed in [9, 13] (see Table 2). SVM which uses motion and area features yielded the best scores for sensitivity and accuracy with 85.97% and 80.24%, respectively. For medical image analysis with limited and imbalanced datasets, sensitivity and specificity are considered the most important metrics [9]. Our methods outperformed the baseline in terms of specificity by a minimum improvement of 11.75% for Harris-Random Forest, and as much as 24.83% against CNN. Results for accuracy and precision are also close to the baseline.

A major challenge with this dataset is the high-degree of class imbalance where healthy segments has a total of 444 while infarcted segments only has 210. Because of this, we further extend our experiment by creating synthetic data using Synthetic Minority Over-Sampling Technique (SMOTE) [7] on the corner Harris features. The infarcted segment samples for each fold are up-sampled to 350 during training. With more test data, the overall results for Harris-SVM and Harris-RF significantly improved. Harris-SVM has higher sensitivity scores from a base score of 47.16% to an improved score of 73.85%. However, other results are much lower, which is a downside of synthetic up-sampling. The results of this experiment is shown in Table 3.

Finally, in the last row of table 3 we show an initial experimentation using CNN with Long Short-Term Memory (LSTM) networks. LSTM is an artificial recurrent neural network (RNN) which processes sequential data through feedback connections [12]. This network is mainly used for speech recognition and video recognition. The results in table 4 shows that despite its performance for sequential classification tasks, CNN-LSTM relies on adequate and balanced samples.

Note that the classification task is a multi-label problem where a particular echo can consist of different combinations of infarcted and healthy segments. For the entire dataset, there are a handful of echo videos with only one class combination which makes splitting the dataset into equal folds cumbersome. This is one reason why most of our results for MI classification yielded high variance across different folds.

6 Conclusion

In this work we have shown that simple pixel-level augmentations can significantly improve learning and generalizability of segmentation networks. Our method which preprocesses low-quality echocardiograph frames yielded better average scores and lower variance for specificity, precision, F1, and accuracy. Among the implemented enhancement techniques, gray-level slicing had the best results because it removes low pixel intensity values per echo frame.

For the classification task, our methods significantly improved specificity but had much lower scores in sensitivity and F1 scores. Even after applying class balancing and utilizing sequential networks, the handcrafted features in [9] still outperformed our methods. Despite these results, our work is one of the earliest, if not the first, attempt to apply computer vision algorithms and deep learning techniques on the HMC-QU dataset. We hope that our results will serve as benchmarks for future works related to this dataset and motion analysis in medical datasets.

7 Future Works

We recommend the following improvements: 1) An end-to-end implementation of segmentation and classification of myocardial infarction as also states in [9]; 2) Implement advanced computer visions algorithms which can extract features on sequential data; 3) Apply one-shot learning approaches for limited, multi-label problems; and 4) Apply better techniques in data augmentations.

References

[1] Rupal R. Agravat and Mehul S. Raval. 2021. A Survey and Analysis on Automated Glioma Brain Tumor Segmentation and Overall Patient Survival Prediction. *Archives of Computational Methods in Engineering* (Mar 2021). https://doi.org/10.1007/s11831-021-09559-w

[2] Jeffrey L. Anderson and D. Morrow. 2017. Acute Myocardial Infarction. *The New England journal of medicine* 376 21 (2017), 2053–2064.

[3] Abhir Bhandary, Ananth Prabhu, Venkatesan Rajinikanth, Palani Krishnan, Suresh Satapathy, David Robbins, Charles Shasky, Yu-Dong Zhang, Joao Tavares, and N. Raja. 2020. Deep-learning framework to detect lung abnormality – A study with chest X-Ray and lung CT scan images. *Pattern Recognition Letters* 129 (01 2020), 271–278. https://doi.org/10.1016/j.patrec.2019.11.013

[4] G. Bradski. 2000. The OpenCV Library. *Dr. Dobb's Journal of Software Tools* (2000).

[5] Alexander Buslaev, Vladimir I. Iglovikov, Eugene Khvedchenya, Alex Parinov, Mikhail Druzhinin, and Alexandr A. Kalinin. 2020. Albumentations: Fast and Flexible Image Augmentations. *Information* 11, 2 (2020). https://doi.org/10.3390/info11020125

[6] Kumardeep Chaudhary, Olivier B. Poirion, Liangqun Lu, and Lana X. Garmire. 2018. Deep Learning–Based Multi-Omics Integration Robustly Predicts Survival in Liver Cancer. *Clinical Cancer Research* 24, 6 (2018), 1248–1259. https://doi.org/10.1158/1078-0432.CCR-17-0853 arXiv:https://clincancerres.aacrjournals.org/content/24/6/1248.full.pdf

[7] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. 2002. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research* 16 (Jun 2002), 321–357. https://doi.org/10.1613/jair.953

[8] N. Dalal and B. Triggs. 2005. Histograms of oriented gradients for human detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* 1 (2005), 886–893 vol. 1.

[9] Aysen Degerli, Morteza Zabihi, Serkan Kiranyaz, Tahir Hamid, Rashid Mazhar, Ridha Hamila, and Moncef Gabbouj. 2021. Early Detection of Myocardial Infarction in Low-Quality Echocardiography. *IEEE Access* 9 (2021), 34442–34453. https://doi.org/10.1109/access.2021.3059595

[10] Linlin Gao, Haiwei Pan, Qing Li, Xiaoqin Xie, Zhiqiang Zhang, Jinming Han, and Xiao Zhai. 2017. Brain medical image diagnosis based on corners with importance-values. *BMC Bioinformatics* 18 (11 2017). https://doi.org/10.1186/s12859-017-1903-6

[11] C. G. Harris and M. Stephens. 1988. A Combined Corner and Edge Detector. In *Alvey Vision Conference*.

[12] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-term Memory. *Neural computation* 9 (12 1997), 1735–80. https://doi.org/10.1162/neco.1997.9.8.1735

[13] Serkan Kiranyaz, Aysen Degerli, Tahir Hamid, Rashid Mazhar, Rayyan Ahmed, Rayaan Abouhasera, Morteza Zabihi, Junaid Malik, Ridha Hamila, and Moncef Gabbouj. 2020. Left Ventricular Wall Motion Estimation by Active Polynomials for Acute Myocardial Infarction Detection. arXiv:2008.04615 [eess.IV]

[14] Claes Ladefoged, Philip Hasbak, Charlotte Hornnes, Liselotte Højgaard, and Flemming Andersen. 2021. Low-dose PET image noise reduction using deep learning: application to cardiac viability FDG imaging in patients with ischemic heart disease. *Physics in Medicine & Biology* 66 (02 2021). https://doi.org/10.1088/1361-6560/abe225

[15] Fengbei Liu, Yu Tian, Filipe R. Cordeiro, Vasileios Belagiannis, Ian Reid, and Gustavo Carneiro. 2021. Noisy Label Learning for Large-scale Medical Image Classification. arXiv:2103.04053 [cs.CV]

[16] Ying Liu and Feixiao Long. 2019. Acute lymphoblastic leukemia cells image analysis with deep bagging ensemble learning. *bioRxiv* (2019). https://doi.org/10.1101/580852 arXiv:https://www.biorxiv.org/content/early/2019/03/21/580852.full.pdf

[17] Francisco Lopez, Andres Varela, Oscar Hinojosa, Mauricio Mendez, Dinh-Hoan Trinh, Jonathan ElBeze, Jacques Hubert, Vincent Estrade, Miguel Gonzalez, Gilberto Ochoa, and Christian Daul. 2021. Assessing deep learning methods for the identification of kidney stones in endoscopic images. arXiv:2103.01146 [eess.IV]

[18] Ana Martinez-Naharro, Philip Hawkins, and Marianna Fontana. 2018. Cardiac amyloidosis. *Clinical Medicine* 18 (04 2018), s30–s35. https://doi.org/10.7861/clinmedicine.18-2-s30

[19] Dania Mohty, Thibaud Damy, Pierre Cosnay, Najmeddine Echahidi, Danielle Casset-Senon, Patrice Virot, and Arnaud Jaccard. 2013. Cardiac amyloidosis: Updates in diagnosis and management. *Archives of Cardiovascular Diseases* 106, 10 (2013), 528–540. https://doi.org/10.1016/j.acvd.2013.06.051

[20] World Health Organization. 2021. Cardiovascular diseases (CVDs). (2021). https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

[22] Joseph Rahman, Emelie Helou, Ramona Gelzer-Bell, Richard Thompson, Chih Kuo, E. Rodriguez, Joshua Hare, Kenneth Baughman, and Edward Kasper. 2004. Noninvasive Diagnosis of Biopsy-Proven Cardiac Amyloidosis. *Journal of the American College of Cardiology* 43 (02 2004), 410–5. https://doi.org/10.1016/j.jacc.2003.08.043

[23] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:1505.04597 [cs.CV]

[24] Ahmed Serag, Gillian Macnaught, Fiona Denison, Rebecca Reynolds, Scott Semple, and James Boardman. 2017. Histograms of Oriented 3D Gradients for Fully Automated Fetal Brain Localization and Robust Motion Correction in 3 T Magnetic Resonance Images. *BioMed Research International* 2017 (01 2017). https://doi.org/10.1155/2017/3956363

[25] Patrice Simard, David Steinkraus, and John Platt. 2003. Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis. 958–962. https://doi.org/10.1109/ICDAR.2003.1227801

[26] Tanawut Tantimongcolwat, Thanakorn Naenna, Chartchalerm Isarankura-Na-Ayudhya, M. Embrechts, and Virapong Prachayasittikul. 2008. Identification of ischemic heart disease via machine learning analysis on magnetocardiograms. *Computers in biology and medicine* 38 (08 2008), 817–25. https://doi.org/10.1016/j.compbiomed.2008.04.009

[27] Kristian Thygesen, Joseph S. Alpert, Harvey D. White, and null null. 2007. Universal Definition of Myocardial Infarction. *Journal of the American College of Cardiology* 50, 22 (2007), 2173–2195. https://doi.org/10.1016/j.jacc.2007.09.011 arXiv:https://www.jacc.org/doi/pdf/10.1016/j.jacc.2007.09.011

[28] Stéfan van der Walt, Johannes L. Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D. Warner, Neil Yager, Emmanuelle Gouillart, Tony Yu, and the scikit-image contributors. 2014. scikit-image: image processing in Python. *PeerJ* 2 (6 2014), e453. https://doi.org/10.7717/peerj.453

[29] Guotai Wang, Wenqi Li, Maria A. Zuluaga, Rosalind Pratt, Premal A. Patel, Michael Aertsen, Tom Doel, Anna L. David, Jan Deprest, Sebastien Ourselin, and et al. 2018. Interactive Medical Image Segmentation Using Deep Learning With Image-Specific Fine Tuning. *IEEE Transactions on Medical Imaging* 37, 7 (Jul 2018), 1562–1573. https://doi.org/10.1109/tmi.2018.2791721

[30] Harvey D White and Derek P Chew. 2008. Acute myocardial infarction. *The Lancet* 372, 9638 (2008), 570–584. https://doi.org/10.1016/S0140-6736(08)61237-4

[31] Yiwen Xu, Ahmed Hosny, Roman Zeleznik, Chintan Parmar, Thibaud Coroller, Idalid Franco, Raymond H. Mak, and Hugo J.W.L. Aerts. 2019. Deep Learning Predicts Lung Cancer Treatment Response from Serial Medical Imaging. *Clinical Cancer Research* 25, 11 (2019), 3266–3275. https://doi.org/10.1158/1078-0432.CCR-18-2495 arXiv:https://clincancerres.aacrjournals.org/content/25/11/3266.full.pdf

[32] Kuo Yang, Emad A. Mohammed, and Behrouz H. Far. 2021. Detection of Alzheimer's Disease Using Graph-Regularized Convolutional Neural Network Based on Structural Similarity Learning of Brain Magnetic Resonance Images. arXiv:2102.13517 [cs.CV]

[33] Hang Zhang, Chongruo Wu, Zhongyue Zhang, Yi Zhu, Haibin Lin, Zhi Zhang, Yue Sun, Tong He, Jonas Mueller, R. Manmatha, Mu Li, and Alexander Smola. 2020. ResNeSt: Split-Attention Networks. arXiv:2004.08955 [cs.CV]

[34] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. 2018. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. arXiv:1807.10165 [cs.CV]