UNIVERSITY OF CALIFORNIA,
IRVINE

Deep learning-based framework for cardiac function assessment in embryonic zebrafish from heart beating videos

THESIS

submitted in partial satisfaction of the requirements
for the degree of

MASTER OF SCIENCE

in Electrical and Computer Engineering

by

Amir mohammad Naderi

Thesis Committee:
Professor Hung Cao, Chair
Professor Yanning Shen
Professor Zhou Li

2023

DEDICATION

To

my parents and friends

in recognition of their worth

TABLE OF CONTENTS

Page

LIST OF FIGURES v

ACKNOWLEDGEMENTS vi

ABSTRACT OF THE THESIS vii

INTRODUCTION 1

CHAPTER 1: Overview 4
- Different image processing methods 4
- Background subtraction 5
- High pass filter 6
- Thresholding 6
- Histogram equalization 8
- Edge detection 9
- Color filtering 10
- Machine learning 11

CHAPTER 2: ZACAF model and its method 16
- Semantic image segmentation validation metrics 16
- Literature review 17
- Unet architecture 19
- Experimental animals 20
- Video imaging of beating zebrafish hearts at the embryonic stage 21
- Titin truncated Mutants 23
- Dataset 24
- Preprocessing 24
- Quantification of the diameters of the predicted ventricle 24
- Graphical User Interface (GUI) 25
- Assessment of the accuracy of the framework with the defined metrics 26
- Assessment of the performance of the framework for EF 27

CHAPTER 3: Expanding the work to new datasets 30
- Considerations for zebrafish age and use of anesthesia 31
- Considerations for image acquisition 32
- Dataset 33
- Data augmentation 34
- Transfer learning 35
- Test Time Augmentation 35
- Assessment of the EF in Nrap deficient zebrafish 37
- Assessment of the performance of the framework for EF in the test set 38

CHAPTER 4: Discussions and Conclusions 40
Estimations in 2-D videos 40
Consistency of measurement 41
Frame rate issue 42
Mutant fish lines challenges 42
Pear shape of the ventricle in some videos 44
Comparison between EF formulas 44
Discussion on transfer learning 44
Discussion on TTA 45
Conclusion 46

REFERENCES 47

LIST OF FIGURES

Page

Figure 1:. 1
Figure 2:. 14
Figure 3:. 23
Figure 4: 27
Figure 5: 30
Figure 6:. 31
Figure 7: 35
Figure 8: 38
Figure 9: 39
Figure 10: 40
Figure 11:. 41
Figure 12: 44

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my committee chair and principal investigator, Professor Hung Cao, for his unwavering support and guidance throughout my graduate studies in electrical engineering at the University of California, Irvine. Professor Cao has provided a comfortable environment for me to conduct my research, and his insightful advice has been invaluable in helping me navigate through difficult situations.

I would also like to thank my committee members, Professor Yanning Shen and Professor Zhou Li, for their valuable contributions to my research. Their expertise and constructive feedback have been instrumental in shaping the direction of my work.

Moreover, I am grateful to the Department of Electrical Engineering and Computer Science at University of California Irvine for providing me with a world-class education and research opportunities. I also extend my appreciation to the Department of Education for their financial support of my research through Graduate Assistance in Areas of National Need fellowship.

Finally, I would like to acknowledge my family and friends for their unwavering support and encouragement throughout my academic journey. Their love and encouragement have kept me motivated and inspired to achieve my goals.

ABSTRACT OF THE THESIS

Deep learning-based framework for cardiac function assessment in embryonic zebrafish from heart beating videos

by

Amir mohammad Naderi

Master of Science in Electrical and Computer Engineering

University of California, Irvine, 2023

Professor Hung Cao, Chair

In this thesis, application of imaging in assessment of zebrafish (zf) cardiology is discussed. Medical imaging seeks to expose internal structures, as well as to diagnose and treat disease. Medical imaging also establishes a database of normal anatomy and physiology to make it possible to identify abnormalities. Considering that the embryonic zebrafish is transparent, bright field microscopic videos could reveal heart mechanism and could be useful for quantification of it, although microscopic imaging can be useful for adult zebrafish as well. Different imaging methods used in other works is discussed first. Later, the cardiovascular parameters that can be measured using imaging are defined. We then compare different digital image processing and deep learning algorithms that have been employed to process or segment images from zebrafish. At the end of the chapter challenges mutant type are investigated.

INTRODUCTION

By looking at a bright field microscopic image from a zf embryo so much information can be conceived. In figure 1 a microscopic image from a zf can be seen.

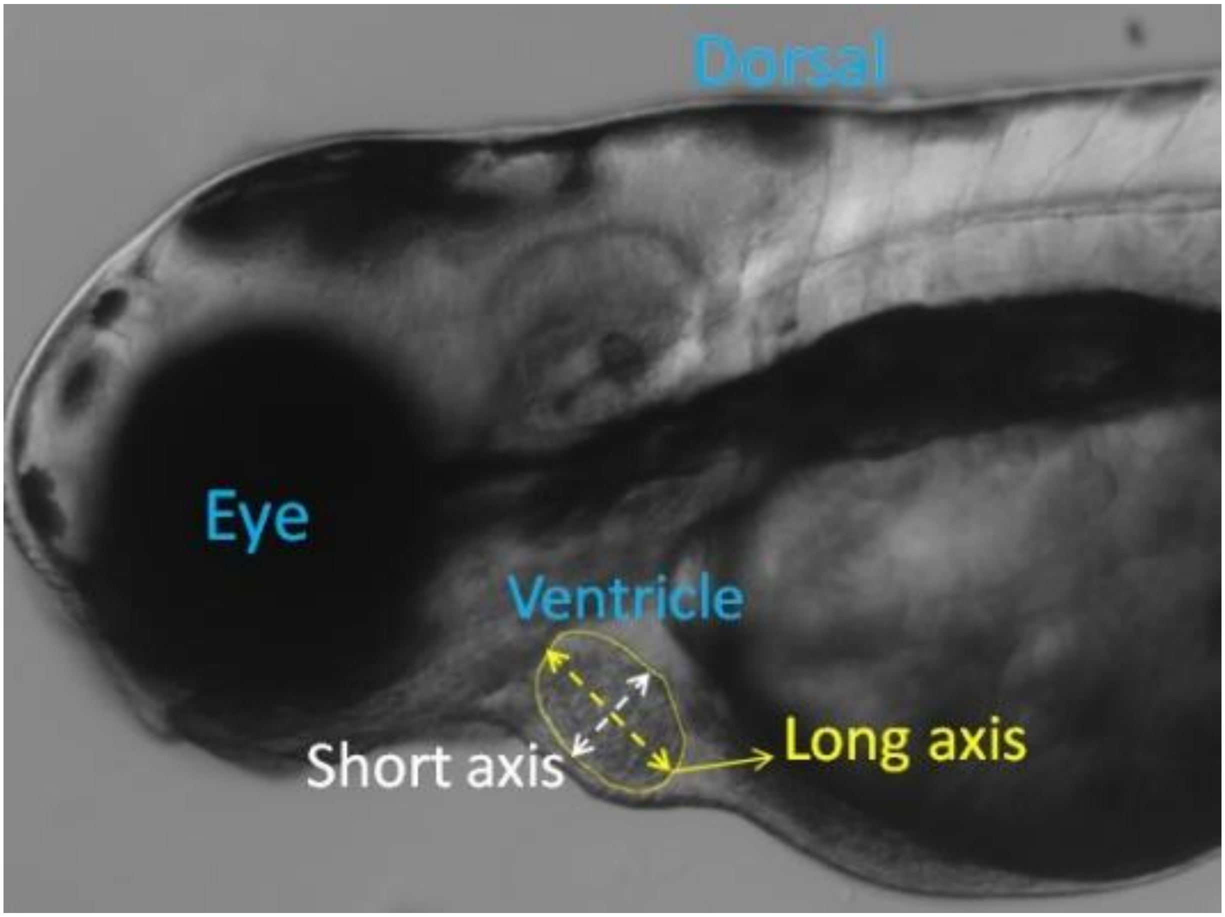

Figure 1:A frame in a video recorded from a 3-dpf zebrafish with segmentation for ventricle border and long and short axes.

For Zebrafish embryos, blood flow velocities are measured to determine cardiovascular function. This can be accomplished simply by following the motions of red blood cells (RBCs) throughout the embryo's body, which are clearly visible due to the embryo's transparent skin. RBC motions (and thus blood flow) can have their acceleration, deceleration, and peak velocity quantified for analysis. RBC motions in the dorsal aorta and the cardinal vein, two major blood arteries in the body, can be observed for this purpose. Individual cell locations are obtained using consecutive frames, and RBC velocity is computed using the coordinates of the cell's location and the time interval between frames as follows:

$$RBC\ velocity = \frac{\sqrt{(x_2-x_1)^2-(y_2-y_1)^2}}{\Delta t} \qquad (1)$$

The embryonic zebrafish (up to three days post-fertilization - dpf) are transparent and have good visibility of their internal organs, including the heart and blood circulation. At this stage, bright field microscopic videos can be used to quantify heart mechanism and morphology. Typically, two-dimensional (2D) movies are recorded for cardiovascular analysis. Then, during the cardiac cycle, continual changes in ventricular wall position would be tracked by first selecting a linear region of interest for the ventricle's borders.

Measurement of myocardial thickness, for example, is critical for determining the magnitude of an induced defect in hypertrophic cardiomyopathy. In Zebrafish embryos, fractional area change (FAC) is a well-established ventricular function metric for evaluating contractility. It can be estimated using 2D still frames of the ventricle at end-diastole (ED) and end-systole (ES). The fully dilated ventricle is designated as ED, whereas the fully contracted ventricle is designated as ES. At these two positions, the ventricular areas (EDA and ESA) are determined, and the FAC is derived as follows:

$$FAC = \frac{(EDV-ESV)}{EDV} \times 100 \qquad (2)$$

Another measure of ventricular contractility is fractional shortening (FS), which can be calculated using the ventricular diameters at ED and ES (Dd and Ds) as follows:

$$FS = \frac{(D_d-D_S)}{D_d} \qquad (3)$$

To determine stroke volume, ejection fraction, and cardiac output, ventricular volumes must be computed. The long- and short-axis diameters (DL and DS) are initially measured from 2D still images. The following volume formula can be employed if the ventricle has a prolate spheroidal shape:

$$Valume = \frac{1}{6} \times \pi \times D_L \times {D_S}^2 \qquad (4)$$

However, if we consider that the shape of the ventricle is unknown while having the 2D shape of the ventricle, the volume can be calculated using the formula bellow:

$$Volume = \frac{8}{3\pi D_L} \times A^2 \text{ (5)}$$

In this formula A is the area of the segmented 2D ventricle [1].

The blood volume pumped from the ventricle for each beat is called stroke volume (SV), and it is easily determined using the ventricle volumes at ED (EDV) and ES (ESV):

$$SV = (EDV - ESV) \qquad (6)$$

The fraction of blood evacuated from the ventricle with each heartbeat is known as ejection fraction (EF), and it may be computed using the formula:

$$EF(\%) = \frac{(EDV - ESV)}{EDV} \times 100 = \frac{SV}{EDV} \times 100 \qquad (7)$$

The following formula can be used to compute cardiac output (CO) from SV and heart rate (HR):

$$CO \left(\frac{nanoliter}{min}\right) = SV \times HR \qquad (8)$$

The time between two identical subsequent points (i.e., ED or ES) in the captured images is used to calculate HR.

Chapter 1: Overview

1.1. Different image processing methods

For calculating HR of the zebrafish from videos, many different methods have been discussed in the literature. To name a few utilizing frequency transforms like fast Fourier transform, filtering, and pixel intensity changes are all employed for quantification of HR and heart rate variability. These methods can be grouped into three categories: Time domain, frequency domain and Blind Source separation. Ling et al, has a review paper on Quantitative measurements of zebrafish heartrate and heart rate variability.[2] However, most of the methods for HR measurement cannot measure the other cardiovascular like heart contractibility measures like EF and FS. On the other hand, most methods used for quantification of heart contractibility can be used to measure HR. Hence, those methods are more general and here we focus on them.

The general idea for quantification of ventricular function metrics for evaluating contractility is sematic segmentation of the heart and more specifically ventricle. By segmenting the ventricle in a series of consecutive frames, ES and ED frames can be found and after that the metrics discussed earlier can all be calculated easily. In colored microscopic videos, segmentation of the heart can be done much easier by filtering red color as red has more intensity in the heart region. In black and white recordings more complicated approaches are needed. At the first look at a beating heart video, it can be concluded that from identification of the borders and periodic movement of the heart are two features that can help for manual segmentation of the ventricle. For movement features, background subtraction can be used and for border features a variety of methods for enhancing and automatic segmentation can be used.

1.2. Background subtraction

Background subtraction is a common technique for dividing the moving parts of a scene in a static camera by segmenting it into background and foreground. Continuous frames from a video are subtracted from each other to locate moving objects in background subtraction. Static pixels can be eliminated because most of the fish body is motionless and the only pixels moving in the video are blood cells and the heart. Frame difference, Gaussian mixture model, kernel density estimation, and codebook are just a few examples of background subtraction approaches. All the methods have varying degrees of accuracy. In a video which the fish is completely static and minimal noise is present this method could be helpful in segmentation of the ventricle. Evidently, other dynamic features like blood cells and vessels, gale movement due to respiration and noisy pixels will be detected in the output of this method but they can be filtered using different approaches. Large moving objects detected that are not a part of ventricle can be removed using specifying a region of interest (ROI) and thresholding the size of the object. Small, detected particles and noise can be removed by filtering. For example, arithmetic mean filter can be used for smoothening and geometric mean filter can be used for removing salt and paper noise. Morphological filters are useful for smoothing binary images, especially for removing small structures and border detection. The morphological filter's concept is a shrink and let grow procedure. The term "shrink" refers to the use of a median filter to round off large structures and remove small structures, with the surviving structures being grown back by the same amount during the grow process[3]. In zebrafish videos, employing the background subtraction often times leads to detecting the ventricle as a region containing a group of multiple objects that form the shape which represents ROI. Applying morphological filters can be helpful because we need to have a single object to represent the ventricle. Nevertheless, background subtraction can result in inaccurate results that cannot be guaranteed.

On the other hand, if we want to detect borders there are several algorithms that can enhance or detect the edges. Enhancing the image in a way that ventricular border can be more visible can be beneficial to researchers for manual and automatic segmentation.

1.3. High pass filter

High pass filters are a form of taking derivative from the time domain signal. Taking derivatives from an image will amplify rapid changes like edges which are why they are referred to as sharpening filters. Laplacian and Sobel filters are some examples of high pass filters used to sharpen an image. Additionally, Gaussian and Butterworth filters can be designed to be high pass filters.

1.4. Thresholding

One of the most basic approaches of image segmentation is Histogram based thresholding. Thresholding can be used to create binary pictures from a grayscale image. In the most basic form of this method each pixel's intensity will be changed to black if it's less than the specified constant threshold or white if it's more than it. There are several different histogram thresholding models, each of which uses a different approach to determine the threshold value.

A. Global thresholding:

Global thresholding involves selecting a threshold value that separates the foreground and background regions of the entire image. This approach assumes that the image has a bimodal histogram, which means that the pixel intensities can be divided into two distinct groups corresponding to the foreground and background regions. The threshold value is typically selected using an iterative method, such as Otsu's method, which maximizes the inter-class variance between the two groups.

B. Adaptive thresholding:

Adaptive thresholding is a variation of global thresholding that adjusts the threshold value locally based on the intensity values of the neighboring pixels. This approach is useful for images with non-uniform illumination or shading, as it can adapt to changes in the local intensity values. The threshold value is typically computed for each pixel using a local region, such as a square or circular neighborhood, and can be based on methods such as the mean, median, or Gaussian distribution of the local intensity values.

C. Iterative thresholding:

Iterative thresholding is a method that involves iteratively adjusting the threshold value based on the intensity values of the pixels within the foreground and background regions. This approach can be used to segment images with complex histograms that do not have clear separations between the foreground and background regions. The threshold value is typically initialized using a global or adaptive thresholding method and then iteratively refined based on the mean or median intensity values of the pixels within each region.

D. Edge-based thresholding:

Edge-based thresholding is a method that involves detecting the edges or boundaries between the foreground and background regions and using these edges to determine the threshold value. This approach can be useful for images with complex structures or textures that do not have clear separations between the foreground and background regions. The threshold value is typically selected based on the gradient magnitude or edge strength of the image and can be determined using methods such as the Canny edge detector or the Laplacian of Gaussian filter.

There are algorithms like Otsu that find the best threshold automatically. For example, Otsu finds a threshold that segments the background and foreground classes of the histogram by minimizing intra-class intensity variance.

1.5. Histogram equalization

Now that we mentioned histograms, discussing histogram equalization would be beneficial. This method usually increases the global contrast of an image, especially when the image is represented by a narrow range of intensity values. The intensities can be better spread on the histogram with this change, employing the full range of intensities evenly. This enables locations with low local contrast to obtain an increase in contrast. This is accomplished through histogram equalization, which effectively spreads out the densely crowded intensity values that reduce visual contrast. In zebrafish videos it is very common for the region of interest to be too dark. This could happen due to poor placement of the fish on the microscope or just the thickness or transparency of some tissues compared with the surrounding tissues. Having a dynamic range of contrast will help manual and automated segmentation of the heart. Histogram equalization works well when the distribution of pixel values is similar all through the image. However, the contrast in certain sections will not be appropriately improved if the image contains regions that are much lighter or darker than the rest of the image. In zebrafish videos this is especially true since the background of light sheet microscopy has the highest intensity in the image and there are different sections in the fish with different transparencies. Hence, Adaptive Histogram Equalization (AHE) solves the problem by transforming each pixel with a transformation function consequent to a neighborhood region. Finally, Contrast Limited AHE (CLAHE) is a variant of adaptive histogram equalization, which doesn’t have the issue of over amplification of noise in regular AHE. [4]

1.6. Edge detection

Edge detection is a fundamental technique in image processing that involves identifying the boundaries or edges between objects in an image. The goal of edge detection is to distinguish between areas of an image with relatively uniform intensity or color and those with sharp transitions or gradients. There are several approaches to edge detection, but the most common involves looking for abrupt changes in intensity or color, such as those that occur at object boundaries. These changes can be detected by analyzing the first derivative of the image along different directions or by using filters that highlight high-frequency components of the image.

One of the most widely used filters for edge detection is the Sobel filter, which is a type of convolution filter that estimates the horizontal and vertical gradients of the image. The Sobel filter applies two 3x3 kernels to the image to estimate the derivatives along the x and y axes, respectively. These derivatives can then be combined to estimate the gradient magnitude and direction at each pixel.

Another common approach to edge detection is to use the Canny algorithm, which was introduced by John Canny in 1986. The Canny algorithm involves several steps, including smoothing the image with a Gaussian filter, calculating the gradient magnitude and direction of the smoothed image, performing non-maximum suppression to thin the edges, and applying hysteresis thresholding to connect weak edges to strong edges. The Canny algorithm, which is one of the most prominent edge detection methods, has a multi-stage algorithm to detect a wide range of edges in images[5]. For fully automated heart segmentation in zebrafish microscopic videos, edge detection algorithms like Canny are usually not robust. The most important problem is that in the zebrafish videos have numerous edges and tissues therefore the canny algorithm detects many different edges next to each other. This makes it imposable to tell which edge belongs to the heart. However, edge detection can be used as preprocessing or one of the steps in an automatic segmentation framework.

1.7. Color filtering

Although not all recordings from zebrafish are colored, color filtering can be extremely useful in processing zebrafish videos. Considering that blood is red, the heart and most vessels express red in the colored microscopic videos. This fact can be used as a segmentation method. A threshold can be set for red which assigns any pixel outside of a specified range for red to black and the pixels in the range will be assigned to white. The result is heart, blood vessels and some noise detected in a binary image.

In literature transgenic animals expressing the myocardial-specific fluorescent reporter have been frequently used. These videos have manually induced feature selection that improves the manual and automatic segmentation. However, in fully automated quantification of cardiovascular metrics like EF, segmentation of ventricle is needed whereas a simple color filtering of the color that the heart is highlighted in is going to segment the whole heart. Akerberg, et al proposed a Convolutional Neural Network (CNN) framework that automatically segments the chambers from the videos and calculates the EF [6]. In this paper a CNN architecture has been used to segment the ventricle and atrium individually in a video where the heart is highlighted. In conclusion, color filtering alone can only be employed as a feature selection method to increase the accuracy of segmentation.

1.8. Machine learning

Those methods, namely edge detection, color filtering, and background subtraction, are not robust with different videos since ventricle edges might have multiple shades of gray. Therefore, researchers attempted to use machine learning approaches to build a fully automated framework. First, unsupervised learning segmentation methods like K-means and Gaussian mixture model (GMM) could be discussed. Secondly, supervised deep learning methods are brought.

1.8.1. K-means

Clustering algorithms are unsupervised algorithms that are like classification algorithms but differ in their underlying principles. When you employ clustering algorithms on your dataset, unexpected features like structures, clusters, and groupings can appear that you wouldn't have imagined. The unsupervised K-Means clustering algorithm is used to separate the interest area from the background. Based on the K-centroids, it clusters or partitions the given data into K-clusters or sections. When you have unlabeled data, the algorithm is used (i.e., data without defined categories or groups). The purpose is to locate specific groups based on some form of data similarity, with K being the number of groups. In clustering an image based, K different color ranges get clustered. In the black and white videos, the same concept could be applied to ranges of gray levels. The K-means algorithm is divided into two parts. It determines the k centroid in the first phase and then moves each point to the cluster with the closest centroid to the data point in the second phase. The Euclidean distance is one of the most widely used methods for determining the distance to the nearest centroid. In the second phase, once the grouping is complete, it recalculates the new centroid of each cluster, calculates a new Euclidean distance between each center and each data point based on that centroid, and allocates the points in the cluster with the shortest Euclidean distance. The member objects and centroid of each cluster in the partition define it. The centroid of each cluster is the place at which the sum of the distances between all the items in the cluster is the smallest. So, over all clusters, K-means is an iterative algorithm that minimizes the sum of distances between each item and its cluster centroid. To generate a noise-free image, median filtering is utilized as a noise removal technique. The segmented image may still have some undesired regions or noise after it has been segmented. As a result, the median filter is applied to the segmented image to improve its quality.[7]

1.8.2. Gaussian mixture model

Gaussian-mixture-based segmentation is based on histogram thresholding which is one of the most used approaches for segmenting images. In histogram thresholding, a picture is divided into two regions or classes: target and background, each with its own uni-modal gray level distribution. As a result, the segmentation problem requires selecting on a suitable threshold for partitioning the image into target and background regions. Gaussian mixture-based segmentation is an image processing technique that involves modeling the image pixels as a mixture of Gaussian distributions and then classifying each pixel into different segments based on the underlying Gaussian distribution to which it belongs. The basic idea behind Gaussian mixture-based segmentation is that each segment in an image can be characterized by a probability distribution that reflects the statistical properties of the pixels in that segment. Specifically, the pixel intensities within each segment are assumed to be normally distributed, and the goal of the segmentation algorithm is to estimate the parameters of the Gaussian distributions that best fit the observed data [8]. To accomplish this, the Gaussian mixture-based segmentation algorithm first initializes a set of Gaussian distributions with random parameters and then iteratively refines these parameters to better fit the observed data. During each iteration, the algorithm computes the likelihood that each pixel belongs to each of the Gaussian distributions and then assigns each pixel to the segment corresponding to the Gaussian distribution with the highest likelihood. The parameters of the Gaussian distributions are then updated based on the pixels assigned to each segment. This process is repeated until the parameters of the Gaussian distributions converge to a stable solution. Once the segmentation is complete, each pixel in the image is assigned to a specific segment, which can be used for further analysis or processing. The accuracy of model parameter estimations and how closely the histogram of an image approximates a Gaussian mixture determine the effectiveness of Gaussian-mixed-based segmentation techniques.[9]

Here, the mentioned methods have been implemented not only to compare with the approach using deep learning described in the next section but also to use for preprocessing. All these are shown in Figure 2, a-d panels.

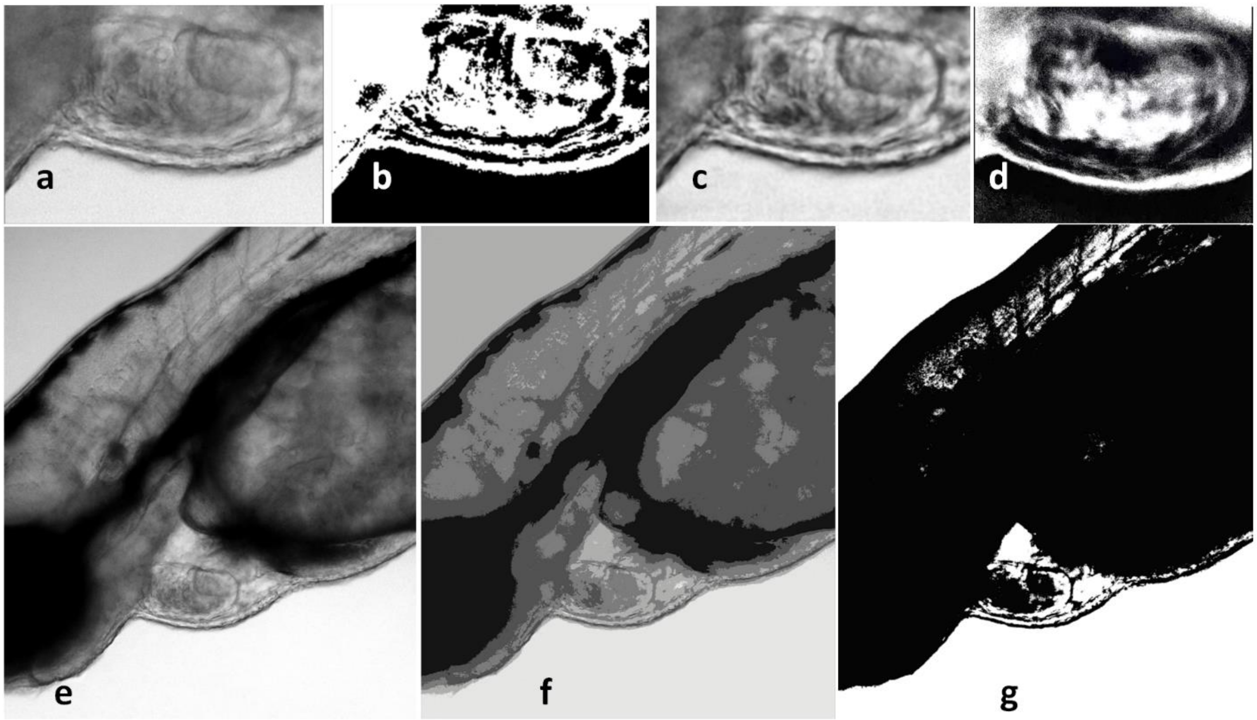

Figure 2: Ventricle segmentation using different methods. Panel a-d: A frame from the video of a 3 dpf zebrafish with 40X zoom undergoing different HBS algorithms. a. Original frame. b. Manual histogram thresholding. c. CLAHE. d. Otsu thresholding. Panel e-g: A frame from the video of a 3 dpf zebrafish with 10X zoom undergoing GMM and K-means approaches. e. Original frame f. GMM. g. K-means.

The abovementioned methods, namely edge detection, color filtering, and background subtraction, are not robust with different videos since ventricle edges might have multiple shades of gray. Therefore, we also attempted to use machine learning approaches to compare. First, unsupervised learning segmentation methods like K-means and Gaussian mixture model (GMM) were applied to the videos. As shown in Figure 2e, f, and g, although these methods improve the visibility of the ventricle borders, the heart's automatic segmentation is not possible. Moreover, much of the unnecessary information (pixels) in the image, particularly in the image generated using K-means, remains.

These methods improve the visibility of the ventricle borders, while the heart's automatic segmentation is not possible. Moreover, much of the unnecessary information (pixels) in the image, particularly in the image generated using K-means, is remaining. However, manual segmentation is extremely tedious work and in most practical research scenarios there are numerus videos recorded and manual segmentation can take be time consuming as well. In conclusion, for a fully automated framework a more robust method is required. For achieving this goal, a few recent papers proposed using deep learning methods.

1.8.3. Semantic image segmentation

Clustering portions of an image that belong to the same object class together is known as semantic segmentation, also known as image segmentation. Due to the categorization of every pixel in an image, it is a type of pixel-level prediction. This prediction of portions can be supervised by having a mask of the portions. For supervised classification of pixels there are many algorithms. There are many methods for image classification like fuzzy measure, decision tree, as well as support vector machine and artificial neural network-based methods. Artificial neural network-based models have been showing great performance and accuracy particularly in biomedical images. The goal of semantic-level image classification is to assign a distinct semantic class to each scene image. The large-scale remote sensing image is manually processed to obtain the scene images. For example, here the object would be the ventricle of the zebrafish. A mask will have two classes: ventricle and the background. To represent the classes in the mask, a specific color scheme is used. The background class is represented by the color black, while the ventricle class is represented by the color white. This choice of colors is important because it aligns with the definition of evaluation metrics such as the Dice coefficient and Intersection over Union (IoU) coefficient.

The Dice coefficient and IoU coefficient are commonly used metrics to assess the performance of image segmentation algorithms or classification models. These metrics compare the overlap between the predicted segmentation or classification result and the ground truth (in this case, the manually created mask). By convention, the background is typically assigned the value of 0 (or black) and the object of interest is assigned the value of 1 (or white) when computing these metrics. By following this convention, the resulting Dice and IoU coefficients can accurately measure the accuracy of the image classification or segmentation algorithm by quantifying the overlap between the predicted and ground truth regions. The higher the coefficients, the better the algorithm performs in capturing the ventricle accurately and discriminating it from the background.

Chapter 2: ZACAF model and its method

2.1. Semantic image segmentation validation metrics

In quantification of cardiovascular metrics from videos using deep learning methods, our objective is to predict the geometrical shape identifying the ventricle with high accuracy in terms of its position, size, and shape with the ground truth. Since the manually created masks are considered as the ground truth, we would expect the predicted shape and the manual mask to be identical or close to them. For validation of the automatic segmentation, we need metrics to show the accuracy of the work. In semantic image segmentation, the most used metrics comprise pixel-wise accuracy, Dice coefficient, and Intersection over Union (IoU).

a. *Pixel-wise Accuracy*

In segmentation of the ventricle, since the mask indicating the ventricle is either white or black, there are only two classes so that we can use the binary case of pixel accuracy. The accuracy is defined as the percent of pixels classified correctly as

$$pixel - wise\ Accuracy = \frac{pixels\ classified\ correctly}{All\ pixels} \qquad (9)$$

In these videos the ventricle has a much smaller area compared with the rest of the frame so this metric alone can be misleading. However, the correct identification of white pixels (which are the pixels creating the background) is essential because they ensure the position, and the shape of the ventricle is also correct.

b. *Dice coefficient*

The dice coefficient is a widely used metric for determining how similar two objects are. It has a scale of 0 to 1, with 1 indicating perfect match or complete overlap. For a binary case, the coefficient is calculated as

$$Dice = \frac{2|(A \cap B)|}{|A|+|B|} \qquad (10)$$

where A is the predicted image and B is the ground truth (manually created mask).

c. *Intersection over union*

It's also known as the Jaccard Index, which is just the area of overlap between the predicted segmentation and the ground truth divided by the area of union between both. This measure runs from 0 to 1, with 0 indicating no overlap and 1 indicating complete overlap. For the binary case, it can be calculated as:

$$J = \frac{|A \cap B|}{|A \cup B|} \qquad (11)$$

2.2. Literature review

To date, most of the reported work only dealt with simple detection of heart rate, such as via edge tracing [10]. Nasrat *et al.* presented a semi-automatic quantification of FS in video recordings of zebrafish embryo hearts [11]. Their software provides automated visual information about the ES and ED stages of the heart by displaying corresponding-colored lines into a motion-mode display. However, the ventricle diameters in frames of ES and ED stages are marked manually, and then the FS is calculated. This will be extremely tedious, time-consuming, and inconsistent when segmentation is done manually for several frames. Akerberg et al. proposed a convolutional Neural Network (CNN) with encoder-decoder SegNet architecture for automatically segmenting and calculating the EF from videos using MATLAB environment. A ground-truth reference dataset was created by manually segmenting systole and diastole for both chambers, across four animals.[6] Nevertheless, particular transgenic animals expressing the myocardial-specific fluorescent reporter and hi-end fluorescence microscopes were used, which cannot be widely applicable for the research community, especially those without access to transgenic lines or fluorescence microscopes. Additionally, Huang et al. showed that transgenic expression of fluorescence protein could cause dilated cardiomyopathy [12], as high levels of expression of some foreign proteins affect the myocardium. On the other hand, unique transgenic mice expressing the myocardial-specific fluorescent reporter and high-end fluorescence microscopes were used, which are not universally applicable to the scientific community, particularly those who do not have access to transgenic lines or fluorescence microscopes.

For a more Inclusive and available example of a fully automatic cardiovascular segmentation for zf, Naderi et al. proposed a framework using U-net to segment monochromic light sheet microscopy videos. [13] In this framework, after preprocessing using sharpening filter and CLAHE, 50 videos of wild and mutant type zf have been manually segmented to be used for the training dataset. The U-net then was trained and validated using the dataset and the deep learning model showed 99.1% for pixel-wise accuracy, 95.04% for Dice coefficient, and lastly 91.24% and for the IoU. They created a graphical user interface to provide an end-to-end platform so researchers can use it conveniently. The framework inputs raw videos and segments the ventricle in each frame of it. The output for each frame is a binary mask of the ventricle. From there diameters of the ventricle in each frame can be calculated. The frames with the largest and smallest area are going to represent ED and ES respectively. Having the ED and ES frames important cardiovascular parameters namely EF, FS, CO, and SV can be quantified. The EF quantification has been validated using 8 videos that haven't been included in the training set. The averages of absolute errors and standard deviations for the automatically calculated EF of the 8 wild type test videos compared to the expert's manual calculation were 6.13% and 3.68%, respectively.

2.3. Unet architecture

UNet is a type of convolutional neural network (CNN) that was originally developed for biomedical image segmentation tasks but has since been widely adopted in other areas of computer vision as well. Its architecture is designed to address the specific challenges of semantic segmentation, which involves assigning a label to each pixel in an image based on its context and relationships with other pixels.

The UNet architecture consists of two main parts: an encoder network and a decoder network. The encoder network is responsible for extracting high-level features from the input image, while the decoder network uses these features to generate a segmentation map. The encoder and decoder networks are connected by skip connections, which allow information to flow directly between them. The encoder network typically consists of several convolutional layers, each of which applies a set of filters to the input image to extract features at different levels of abstraction. The output of each convolutional layer is then passed through a nonlinear activation function, such as a ReLU, to introduce nonlinearity into the network. In addition to the convolutional layers, the encoder network also includes pooling layers, which downsample the feature maps to reduce their spatial dimensions. This helps to capture increasingly abstract features at each layer of the network, while also reducing the computational complexity of the network. The decoder network is responsible for generating the segmentation map based on the features extracted by the encoder network. It consists of a series of transposed convolutional layers, also known as deconvolutional layers, which upsample the feature maps to their original spatial dimensions. The output of each transposed convolutional layer is also passed through a nonlinear activation function, such as a ReLU, to introduce nonlinearity into the network.

The skip connections in UNet allow information to flow directly from the encoder network to the decoder network, bypassing the intermediate layers of the decoder network. This helps to retain spatial information and improve the accuracy of the segmentation, especially in regions with complex background or foreground objects.

To summarize, UNet is a powerful architecture for image segmentation that is designed to capture high-level features of an image and retain spatial information through the use of skip connections. It has been widely used in biomedical image analysis and has also shown promising results in other areas of computer vision, such as natural language processing and robotics. [14] A similar architecture has been employed by Decourt *et al.* to segment the human left ventricle from magnetic resonance imaging (MRI) images [15]. The main idea of the U-net is to complement a traditional contracting network by successive layers, where pooling operations are replaced by up-sampling operators. Besides, a subsequent convolutional layer can then be trained to assemble a precise output based on this information. The training of the network uses the original image as an input and the mask of the corresponding image as the output, and the objective is to minimize the error of the estimation and the mask. In the next part ZACAF details will be discussed.

2.4. Experimental animals

Zebrafish (Danio rerio; WIK strain) were maintained under a 14 h light/10 h dark cycle at 28.5°C. All animal study procedures were performed in accordance with the Guide for the Care and Use of Laboratory Animals published by the U.S. National Institutes of Health (NIH Publication No. 85-23, revised 1996). Animal study protocols were approved by the Mayo Clinic Institutional Animal Care and Use Committee (IACUC #A00002783-17-R20).

2.5. Video imaging of beating zebrafish hearts at the embryonic stage

Zebrafish in the embryonic stages were anesthetized using 0.02% buffered tricaine methane sulfonate (MS222 or Tricaine) (Ferndale, Washington, US) for 2 minutes and then placed lateral side up with the heart facing the lower-left corner. The specimens were held in a chamber with 3% methylcellulose (Thermo Fisher Scientific, Massachusetts, US). The videos were recorded using a Zeiss Axioplan 2 microscope (Carl Zeiss, Oberkochen, Germany) with a 10X lens and differential interference contrast (DIC) capacity. The used Zeiss' Axiocam 702 mono Digital Camera 426560-9010-000 records videos with 60 fps; however, using the Zeiss computer software, videos get stored in 5 fps, 10 fps, and 20 fps. Video clips were processed using ImageJ for manual quantification of cardiac functional indices, including heart rate and fraction shortening, as detailed in the following sections.

2.6. ZACAF

Figure 3 illustrates the architecture of the proposed U-net model with details.

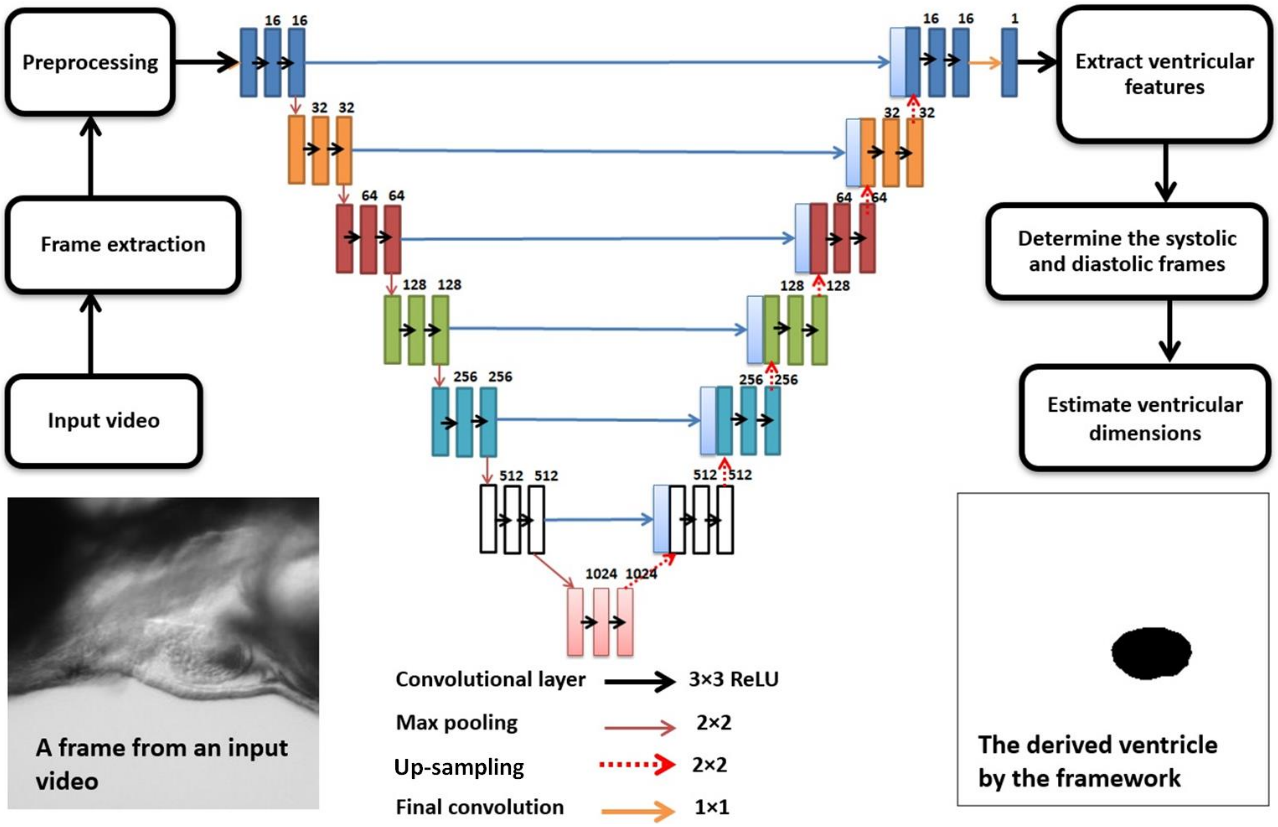
Preprocessing
Frame extraction
Input video
16 16
32 32
64 64
128 128
256 256
512 512
1024 1024
1
Extract ventricular features
Determine the systolic and diastolic frames
Estimate ventricular dimensions
A frame from an input video
The derived ventricle by the framework
Convolutional layer
3×3 ReLU
Max pooling
2×2
Up-sampling
2×2
Final convolution
1×1

Figure 3: The process flow and the U-net architecture. Each rectangle represents a layer and the number above it shows the number of neurons inside. A trained model can estimate a mask of the ventricle from all the extracted frames of the input video. When all the frames have a predicted mask, by determination of ES and ED frames, important cardiac indices like EF, FS, and stroke volume can be automatically calculated and saved in a desired format.

The network consists of a contracting path and an expansive path, which gives it a U-shaped architecture.

The contracting path is a typical convolutional network that consists of repeated convolutions, each followed by a rectified linear unit (ReLU) and a max-pooling operation. Dropouts have been used to prevent overfitting. The architecture has been optimized to obtain the best result. For training, NVidia's T4 GPU from Google Collaboratory was employed. The most commonly used loss functions for semantic image segmentation were deployed to evaluate the model, namely Binary Cross-Entropy and Dice loss function. Cross-entropy can be defined as a measure of the difference between two probability distributions for a given random variable or set of events. It is extensively used for classification problems, and since segmentation is the classification at a pixel level, cross-entropy has been widely used. Binary Cross-Entropy is defined as:

$$Loss_{BCE}(y, \hat{y}) = -(y\log(\hat{y}) + (1 - \mathrm{y})\log(1 - \hat{y})) \qquad (12)$$

where y is the true value and $\hat{y}$ is the predicted outcome.

The Dice coefficient is a commonly used metric in computer vision problems for calculating the similarity between two images. In 2016, it was also adapted as a loss function, namely Dice Loss [16].

$$Loss_{Dice}(y, \hat{y}) = 1 - \frac{2y\hat{y}+1}{y+\hat{y}+1} \qquad (13)$$

The U-net model has been trained with both models, and the performance has been assessed using validation and test sets. Further, the calculation of EF has also been evaluated using both loss functions.

2.7. Titin truncated Mutants

Dilated cardiomyopathy (DCM) is a hereditary, progressive disease, which eventually leads to heart failure [17]. Thus, it is essential to evaluate the early cardiac functions associated with DCM. Dozens of pathogenic genes have been found in the genetic studies of cardiomyopathy, and the incidence rate of DCM is about 1/250 [18]. Titin truncated variants (TTNtv) are the most common genetic factor in DCM, accounting for 25% of DCM cases [19]. Therefore, we have recently restated the allelic heterogeneity in zebrafish segments and established a stable mutation system to assess mutant zebrafish's cardiac functions systematically and accurately. In order to study the mechanobiology of induced defects of these disease models, heart functions need to be reliably evaluated [20].

2.8. Dataset

A training dataset was created employing raw microscopic videos of zebrafish containing 800 pixel-wise annotated images. 50 videos of the lateral view from 50 different 3-dpf zebrafish were analyzed for creating the dataset. 10 of these videos are from the TTNtv mutant line. From each video, 10 to 30 frames were extracted. A total number of 850 frames were extracted for the training set. Each training set has a frame from the video, and a mask manually created showing only the ventricle with ImageJ software. After making the masks, all image and mask sets have been organized into folders. Each set has two folders inside, one for the original extracted frame and the other for its corresponding mask. Finally, all sets were shuffled to avoid overfitting. The validation set with the 10% of the data's size has been split from the dataset before training.

2.9. Preprocessing

In the preprocessing stage, a region of interest is defined, knowing all recordings have the same positioning for the zebrafish. Although this cropping improves the accuracy by removing unnecessary information, it can be avoided to make the framework robust to different video types. Additionally, a sharpening filter accomplished by performing a convolution between a custom weighed kernel and an image is used to make edges more visible. After training, the U-net architecture was able to predict the ventricle segment. The model has been trained several times by applying the mentioned image processing methods to the training images. The method with the best results was CLAHE thresholding which was added to the preprocessing section.

2.10. Quantification of the diameters of the predicted ventricle

The ventricle's diameters are measured for all extracted frames automatically with the contour tool from OpenCV (an open-source computer vision library). The maximum and minimum measured areas of the ventricle in different frames show the ES and ED stages, respectively. Using the measurement of ES and ED frames, we can calculate the ejection fraction (EF), fractional shortening (FS), and stroke volume (SV). Also, the time between two ES (or ED) frames could be used to derive heart rate (HR). The predicted ventricle is assumed to be an ellipsoid. For quantification of EF, the ventricle area can be used (**Eq (5)**) by counting the pixels inside the predicted shape. Since the frames are 2D, we are estimating the ventricle volume to its area. For FS, measurements of the short axis in ES and ED frames are needed. As the ventricle is not a perfect ellipsoid, estimation of the short and long axes can be carried out in two different ways. In the first method, an ellipsoid could be fitted in the predicted shape, and then the axis of the fitted ellipsoid would be measured. The second way is to find the longest line as the long axis of the estimated ellipsoid, which could be found in the geometrical shape; then, the short axis of the ellipsoid is the short axis of the ventricle. In this framework, the 2-D area of the ventricle directly measured from the mask has been used for EF since it is more accurate.

2.11. Graphical User Interface (GUI)

This framework was developed in Python, and thus, for researchers who are not familiar with programming, working with it can be challenging. To address this, a Graphical User

Interface (GUI) has been designed to provide a user-friendly interface to facilitate researchers' process. Moreover, after training the U-net, the trained model can be saved, which means the most computationally heavy part could be done only once. The GUI saves the output files in the CSV format, along with information about EF, FS, diameter readings of the area, short and long axis, and frame numbers. Therefore, each video's data can be easily accessed at anytime and anywhere with the expandable cloud feature. Our ZACAF provides an end-to-end interface to researchers to automatically calculate, classify, and record various cardiac function indices reliably. ZACAF can work with multiple videos simultaneously and output the results in a fraction of the time compared to that of manual segmentation. The deep learning model in the ZACAF can easily be updated and optimized with a new model and data.

2.12. Assessment of the accuracy of the framework with the defined metrics

The model's performance can be seen in **Figure 4**. The model has been trained with two loss functions discussed in section **2.6,** and the best results with parameter tuning are illustrated. The metrics mentioned above resulted in 99.1% for pixel-wise accuracy, 95.04% for Dice coefficient, and lastly 91.24% and for the IoU.

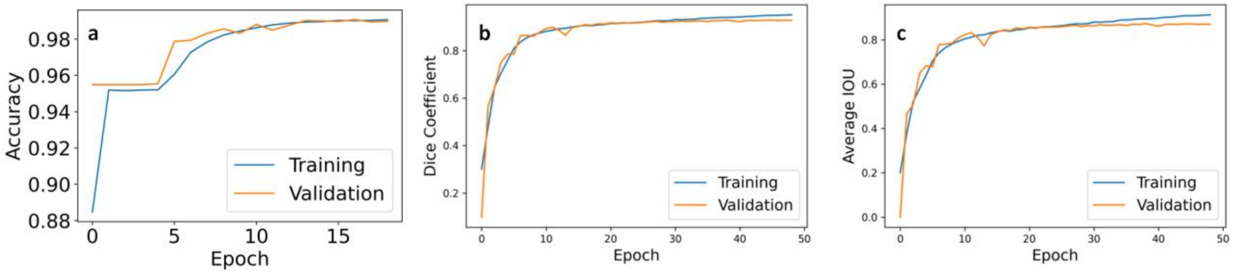

Figure 4: ***The proposed model's performance plotted with the metrics commonly used in semantic image segmentation. a.*** *Pixel-wise accuracy* ***b.*** *Dice coefficient* ***c.*** *IoU metric. This plot shows the performance of the framework with the training and validation sets during the process of training of the deep learning model.*

All mentioned metrics are evaluating the best performing model that had a Dice loss function with an Adam optimizer and a 0.001 learning rate along with decay steps of 240 and a decay rate of 0.95. The validation split was 10% which means 80 sets. Following the training, we visually assessed the framework's ability to correctly segment ventricular chambers and the periodic pulsating movement of it within series of frames of a test video. This process was used in parameter tuning for the deep learning model.

2.13. Assessment of the performance of the framework for EF

The framework was evaluated by comparing the results obtained by manual assessment of EF from an experienced biologist with those using the software since one of the primary purposes of this framework is EF calculation. In this calculation, finding the area in all frames of a video is important because we want to find the ED and ES areas. Hence, assessment should involve the series of frames in a test video rather than having random images in a validation set. For this reason, we assess the performance of ZACAF with EF calculation. First, 8 videos of wildtype zebrafish embryos and another 8 from TTNtv mutant embryos were used as the framework's input. These videos are the test set and have not been used in training. Second, manual processing and estimation were performed for each video to derive EF by an expert to use as the ground truth. The program saves the predicted ventricle masks for every frame of a video, and the ED and ES frames are simply the frames with a maximum and minimum area of the segmented ventricle, respectively. After automatically finding ES and ED frames, the EF of the fish in the input video would be calculated and saved in a CSV file along with other indices calculated. The averages of absolute errors and standard deviations for the calculated EF of the 8 wild type test videos compared to the expert's manual calculation were 6.13% and 3.68%, respectively.

As ED and ES frames are the most important parameters to quantify cardiovascular indices, we plotted the correlation of the automated and manual measurements (Figure 5 a,b). Moreover, Bland–Altman analysis was then used to assess the agreement in manual and automatic ventricle segmentation. Bland–Altman demonstrates the difference that were measured at the same time plotted against the average of the EF with two methods. Larger differences would specify larger disagreement between the two calculations [21]. From 16 test videos two different sets of ES and ED frames (meaning 4 frames from each video) have been manually and automatically segmented. As it can be seen in Figure 6 c the Bland-Altman has been plotted for all pairs of measurements in the same figure with blue and red dots representing mutant and wild fishes respectively. Mean bias and standard deviation calculated was -25.875± 105.102.

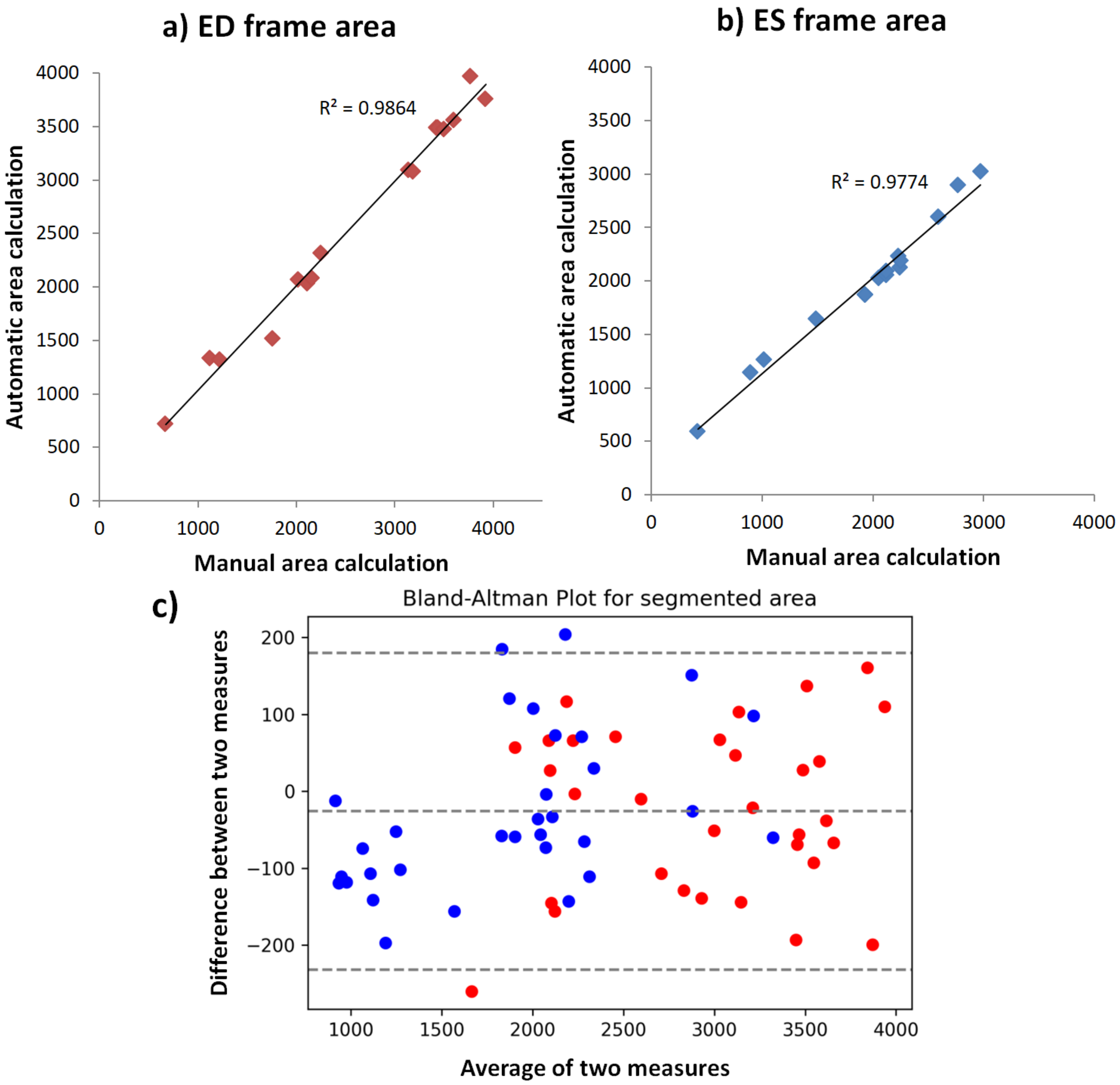

*Figure 5: After finding and measurement of the ventricle area in ED and ES frames of 8 wild type and 8 TTNtv mutant fish with both manual and automated methods, the results are demonstrated in a correlation plot while the calculated EF for the wild and mutant types is plotted in Bland-Altman to demonstrate the agreement of measured values. Linear relation of the measurements with slopes close to 1 shows the accuracy of the ZACAF. (**a**) ED frame area. (**b**) ES frame area. (**c**) Bland-Altman plot for 64 sets of measurements of the segmented ventricle using manual and ZACAF methods. Both mutant and wild have 32 pairs each represented in the plot. Red and blue dots represent wild and mutant fishes respectively. The measurements are in pixels.*

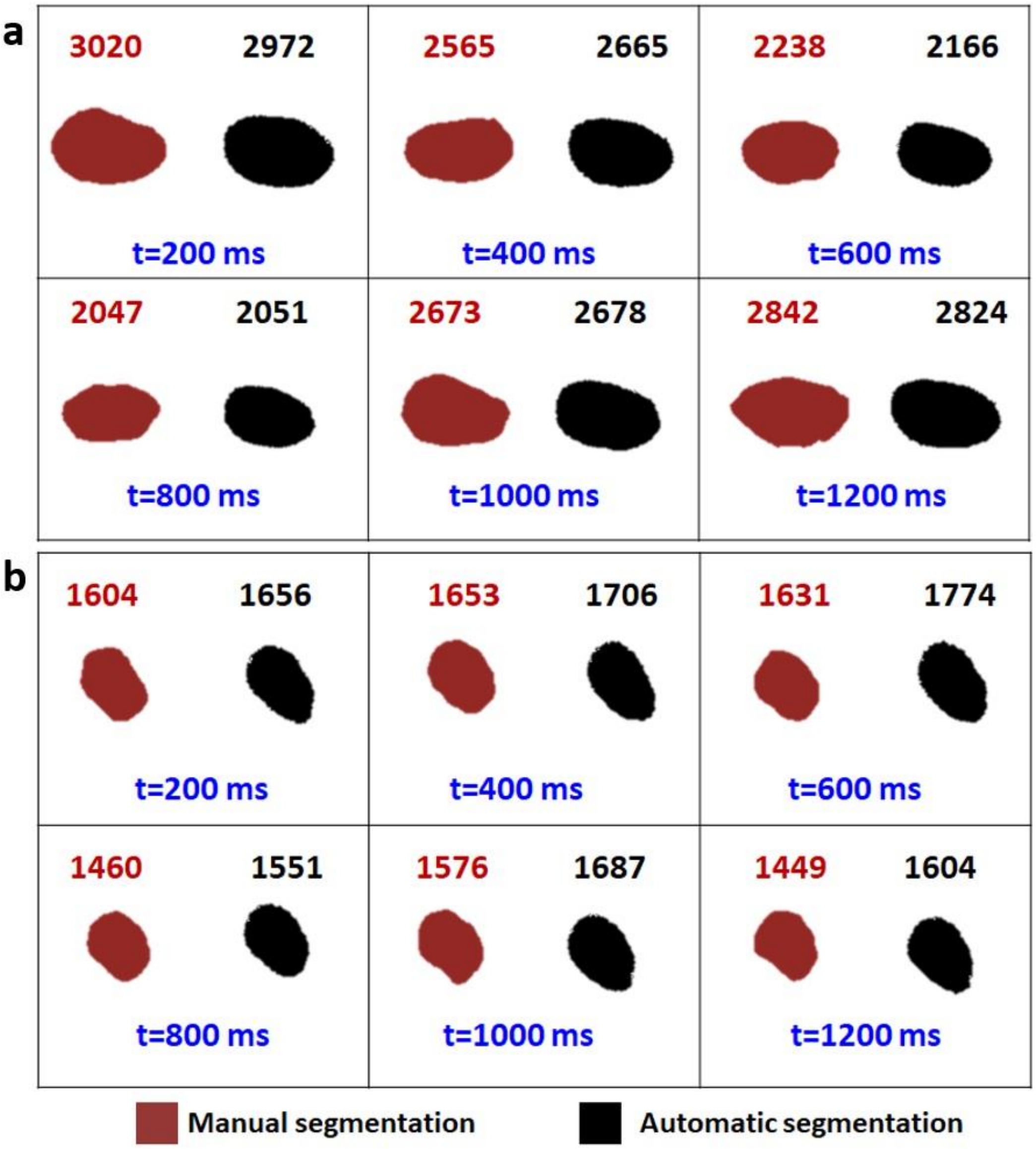

Figure 6: ***Validation of U-net image segmentation framework.*** *The sequential frames from a wild type zebrafish recorded video with fps of 5 are extracted. The respective ventricle mask of each frame is shown in each panel via manual and automatic segmentation. The area of each ventricle is measured and written above its own box. Considering the fps of the videos and the average heart rate of the zebrafish, 6 consecutive frames have been shown in this figure to ensure having at least one full cycle.*

Figure 6 presents the comparison of manual and automatic segmentation of the ventricle in 6 continuous frames to cover an entire cardiac cycle for both wild type (**a)** and TTNtv (**b)**. In manual segmentation, measures were done using the freehand selection tool in the ImageJ software.

Chapter 3: Expanding the work to new datasets

Similar to ZACAF, several image processing frameworks have been proposed to automate the process of automatic quantification of zf cardiovascular parameters. However, most of these works rely on supervised deep learning architectures. However, supervised methods tend to be overfitted on their training dataset. This means that applying the same framework to new data with different imaging set up and mutant types can result in severe decrease of the performance. Here, we take Nrap genotype, and Zebrafish Automatic Cardiovascular Assessment Framework (ZACAF) as an example to demonstrate a modified framework. In this modification we apply data augmentation, Transfer learning, and test time augmentation to ZACAF to improve the general performance and propose a protocol for other researchers to be able to apply the available frameworks for their own data.

Nebulin Related Anchoring Protein (NRAP) is a protein coding gene expressed in cardiac and skeletal muscle. NRAP is a member of the Nebulin family of proteins and promotes myofibril assembly by colocalization with actin during myoblast fusion in early stages of development in skeletal muscle [22]. Previous studies in zebrafish have shown that overexpression of Nrap results in severe skeletal muscle myopathy. Additionally, reducing levels of Nrap in a Klhl41 deficient zebrafish model resulted in a less severe phenotype, as klhl41 is a regulator of NRAP ubiquitination [23]. In cardiac muscle, NRAP plays a role in myofibril assembly and is localized at cardiac intercalated discs [24]. In mouse models of dilated cardiomyopathy, NRAP is overexpressed early in development [24]. Thus, downregulation of NRAP suggests therapeutic advantages in multiple model organisms. Interestingly, a homozygous truncating mutation of *NRAP* was found in a human dilated cardiomyopathy patient. However, the variant was not detected in a cohort of 231 dilated cardiomyopathy patients, and the patient's unaffected brother carries the same mutation, suggesting a low penetrance and allele risk [25]. Due to the potential therapeutic advances of NRAP downregulation found in Zebrafish and Mice, as well as a report of an NRAP truncating variant in a dilated cardiomyopathy patient, we aimed to investigate the cardiac effects, specifically ejection fraction and fractional shortening, of NRAP downregulation in embryonic zebrafish with ZACAF deep learning model. Despite a report of NRAP downregulation associated with cardiomyopathy in a human patient, we see no significant differences in ventricle shape, ejection fraction and fractional shortening between genotypes in embryonic Zebrafish.

3.1. Considerations for zebrafish age and use of anesthesia

The use of anesthesia is necessary for fish greater than 2 days post fertilization, as the fish become mobile at 3dpf as their swim bladder inflates. Tricaine anesthesia, commonly used in zebrafish, may impact the animal's cardiac system. Thus, it is necessary to record the animal no later than 2dpf or deliver anesthesia in a standardized way where each fish receives the same dose. For 5dpf zebrafish, we placed the zebrafish in a separate dish of 16mg/L of Tricaine anesthesia in egg water for exactly 5 minutes prior to transferring the zebrafish to a glass slide emended in 3% methyl cellulose for image acquisition. This method proved to be time consuming, so we opted to record animals at 2dpf instead. At 2dpf, the zebrafish cardiac system is fully developed and functional (5). Additionally, at 2dpf, the zebrafish begin to hatch from their chorion, allowing for their bodies to become straight. It is important that the animal is in this stage before recording, so there is a narrow window for acquisition – the fish must have hatched from the chorion but are not yet mobile. To appropriately time this, dividers should be used at the time of crossing. Additionally, movement of egg water in the dish via transfer pipet will stimulate chorion hatch in the morning and recordings should be made shortly after in the afternoon. An additional consideration besides the use of anesthesia for the age of fish is the pigmentation of the animal. At 2dpf, the animal is transparent, while at 5dpf, skin pigmentation may hinder the ability to acquire a clear image of the ventricle.

3.2. Considerations for image acquisition

Accurate calculations of ejection fraction require a frame of the organism's ventricle in a true systolic and diastolic position. Assuming a resting heart rate of 60 to 100 beats per minute, a high-speed recording camera is necessary. Initially, videos were recorded on a Leica K3 camera (Leica, Germany) with a maximum frame rate of 30 frames per second attached to a Leica S9D microscope at 5x magnification. Videos were processed with the Leica LASX software. However, with this speed, we were unable to visualize the heart in a true diastolic and systolic position. Thus, on the improved setup, we used a FastCam-PCI high-speed digital camera (Photron, USA) with a frame rate of 250fps attached to a Zeiss upright microscope at 10X and processed with the FastCam-PCI image capture board. Additionally, at 5X, we were unable to visualize the boundaries of the ventricle, while at 10X, there is a clear visualization of the ventricle boundaries and red blood cells which can be seen in figure 3. In order to maximize video quality while reducing file size, videos were acquired in greyscale with a resolution of 512 × 480 pixels and recorded for only roughly 4.35 seconds, enough time to capture roughly 8 cardiac cycles and 1,088 frames with a 0.004 shutter speed (5). An additional consideration for recording is that fluorescent lights illuminating the microscope room (or the microscope light bulb itself) have a specific frequency. Fluorescent lights can result in horizontal lines, banding, or flickering in the video, depending on the shutter speed that was utilized to capture the footage. Although it is advised to shoot in varied lighting conditions, changing the camera's shutter speed can also help resolve this problem.

One other important factor that needs to be considered for imaging is the placement of the fish under the microscope. Firstly, in most literature for quantification of cardiac function using 2D imaging the ventricle is assumed to be an ellipsoid. Hence, during image acquisition it is important to make sure the fish is properly positioned to its side. improper placement of the fish under the microscope can result in the ventricle having a pear-shaped structure which can be observed in figure 3 on the left. This will eventually result in inaccuracy in the quantification of EF. Furthermore, it is important to know that the placement of the fish under the microscope is a feature in deep learning models.

Uniform placement of the fish across the videos recorded for the data set can affect the model into only responding accurately to test videos with similar placement. To train a robust framework Data augmentation must be used in the process of training the deep learning model. Using data augmentation will make the framework less prone to placement of the fish.

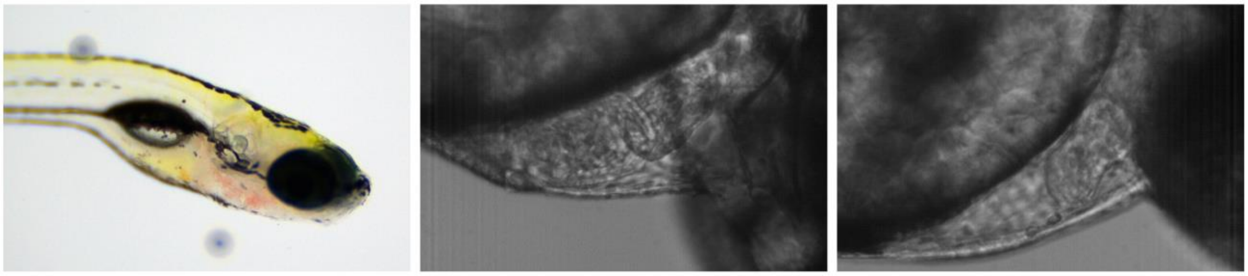

Figure 7: Comparing the visibility of the zf ventricle using 10x and 5x zoom and a pear-shaped ventricle, respectively from left to right. As can be seen, in the 5x zoom image the borders of the ventricle cannot be identified. However, in the 10x image in the middle both chambers can be seen. The rightest image is an example of a pear-shaped ventricle.

3.3. Dataset

In this study, a training dataset was constructed using raw microscopic videos of zebrafish, comprising a total of 410 pixel-wise annotated images. To create this dataset, 41 videos of the lateral view from 41 different 2-dpf zebrafish were analyzed. Specifically, 9 of these videos were obtained from the Nrap mutant line, 19 from the Heterozygous line, and the remaining 9 from the wild type variant. It is noteworthy that during the process of manual segmentation, the experts responsible for the task were provided with videos with randomly generated names which hide the label identifying the fish's genotype. This will make sure that they would not have a bias while performing the segmentation. From each video, 10 sequential frames were extracted, resulting in a total of 410 frames for the training set. Each training set consisted of an original frame extracted from the video and a manually created mask showing only the ventricle, using ImageJ software. Following mask creation, all image and mask sets were organized into folders, with each set containing two folders: one for the extracted original frame and the other for its corresponding mask. For the validation set, two end-systolic (ES) and two end-diastolic (ED) frames were extracted and manually segmented from each video, resulting in a total of 82 images with their corresponding masks. This will make sure that the validation set is independent from the training set.

3.4. Data augmentation

In semantic segmentation, data augmentation is a technique used to increase the size of the training dataset by creating new training examples from the existing ones [26]. This is achieved by applying a range of transformations to the original training images, resulting in new images that are still representative of the same underlying scene or object, but with variations in appearance. Data augmentation is commonly used in deep learning-based computer vision tasks, including semantic segmentation, to prevent overfitting and improve the generalization capability of the model. By augmenting the training data, the model is exposed to more variations in the input data, leading to improved performance on new data. Some common image transformations used for data augmentation in semantic segmentation include horizontal and vertical flipping, rotation, scaling, cropping, and color jitter [27]. These transformations can be applied randomly or systematically during the training process to generate a diverse set of training examples.

In the original ZACAF implementation no augmentation was used since all the videos were recorded with a uniform placement under the camera. However, in this dataset the orientation of the fish was random. This augmentation assures that the framework is less prone to the manner that the fish is placed under the microscope and cameras setup, which gives the user more freedom while recording. Additionally, data augmentation is one of the methods used for improving the performance with limited data and reducing overfitting. Lastly, using data augmentation enables Test Time Augmentation (TTA) which is discussed in the next section. Here only horizontal and vertical flipping transformations have been used to imitate all the possible fish placements during the recording process.

3.5. Transfer learning

Transfer learning is a machine learning technique where a pre-trained model, typically trained on a large dataset, is used as a starting point for a new task or problem [28]. The pre-trained model has already learned a set of feature representations that are applicable to a wide range of problems, and these learned features can be transferred and fine-tuned to the new problem with a smaller dataset. This allows for faster training and better performance than training a model from scratch on the new dataset. If transfer learning is not utilized, the model will be initialized with random weights. In this case, the original ZACAF model was trained on a dataset of zebrafish recorded by a different group using a different microscope setup. The dataset included less mature fish and different mutant types. However, the features learned during the original model's training for ventricle segmentation from zebrafish are expected to improve the training for the new dataset. Therefore, we employed the pre-trained weights from the original model for the new model.

3.6. Test Time Augmentation

Test-time augmentation (TTA) is a technique used in computer vision, including semantic segmentation, to improve the accuracy of models during inference. In semantic segmentation, TTA involves applying various image transformations or augmentations to the test images and feeding them through the model multiple times to obtain an ensemble of predictions. These predictions are then combined, typically by averaging or voting, to obtain a final segmentation map. TTA helps to account for variability in the test data and reduces the risk of overfitting to the training data. By using TTA, a model can be more robust and accurate on previously unseen data. Some common image augmentations used in TTA for semantic segmentation include flipping, rotating, scaling, and cropping. These transformations create multiple versions of the same image, which can be fed through the model to obtain a diverse set of predictions.

It is important to note that TTA can increase the computational cost of inference since the model needs to be run multiple times for each image. However, the benefits in terms of accuracy improvement can often outweigh this cost. Moshkov et al incorporated TTA in the task of semantic segmentation of single-cell analysis of microscopy images based on U-net and Mask R-CNN deep learning models which showed improvement in the prediction accuracy [29]. A set of test time augmentation techniques was then applied to the test image to generate multiple predictions, which were subsequently combined to obtain a final prediction. Specifically, horizontal flipping, vertical flipping, and a combination of both were utilized to create three additional variations of the original test image. For each of these variations, a prediction was obtained using the semantic segmentation model. The four predictions were then combined by taking their element-wise average and thresholding the result with a value of 0.2. An example of TTA applied to a random image from the validation set can be seen in figure 4.

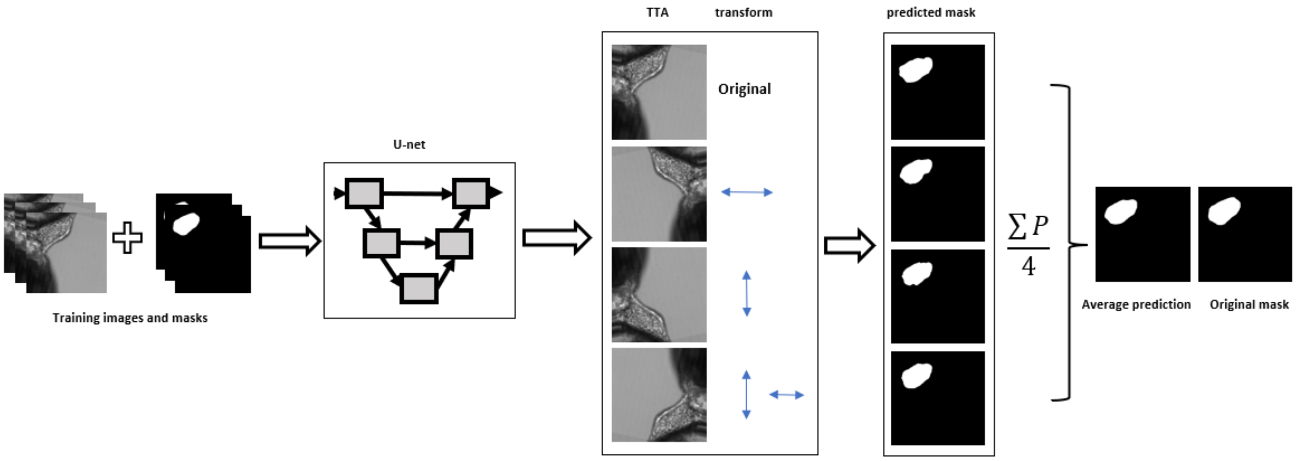

Figure 8:Implementation of the test time augmentation techniques. The U-net architecture is trained by augmented dataset composed of images and their corresponding masks and the TTA will be applied to the test set's output. The transforms used in TTA are horizontal and vertical flipping and their combination. These transformations along with the original prediction make 4 images which then would be subjected to an element-wise average. On the far left of the figure the final prediction resulted from the TTA can be compared with the original manually segmented mask.

3.7. Assessment of the EF in Nrap deficient zebrafish

N=41 Zebrafish were recorded and genotyped via qPCR with TaqMan Custom SNP Genotyping Assays, resulting in 19 heterozygotes (46%), 9 wild type (22%), and 13 mutants (31%). After measurement of EF using the modified ZACAF a one-way ANNOVA test revealed no significant differences between Ejection Fraction and Fractional Shortening in all three genotypes. (Figure 3) Observations of "pear shaped" ventricles were made at the time of image processing (figure 5) and we found no significant correlation. Out of n=41 zebrafish, 11 abnormal ventricles were observed yielding a frequency of 0.27. Within the group of abnormal ventricles, 18% were mutant, 36% were heterozygous and 45% were wild type.

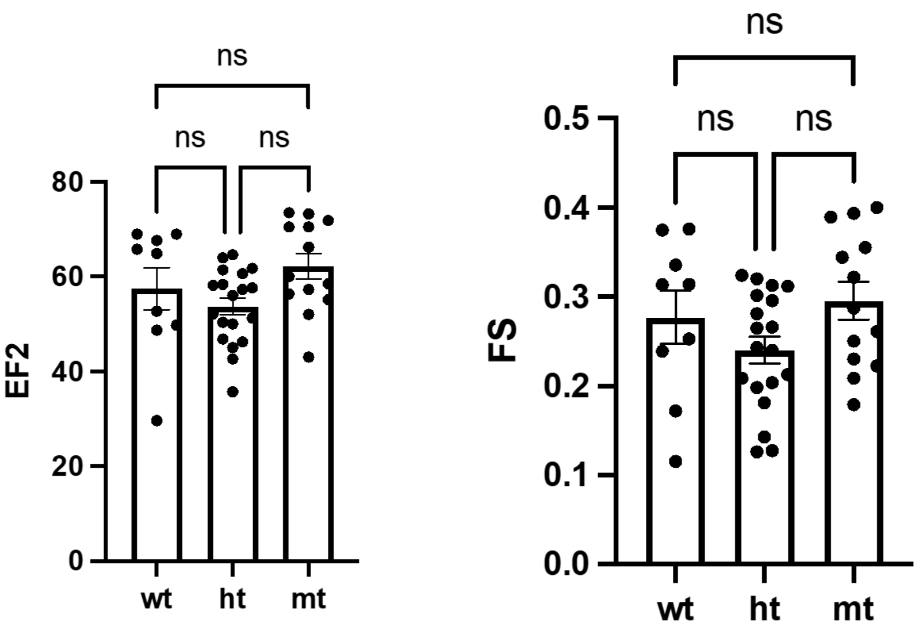

Figure 9: NRAP Zebrafish Model 2dpf Ejection Fraction and Fractional Shortening From 9.29.22 with one-way ANOVA

3.8. Assessment of the performance of the model with the defined metrics

The mentioned metrics assess the performance of the best-performing model, which was trained using a Dice loss function, an Adam optimizer, and a learning rate of 0.001 with decay steps of 240 and a decay rate of 0.95. The validation split consisted of 20%, or 80 sets. The performance of the model is illustrated in figure 6 on the left. For callbacks, a model check pointer was implemented to keep the model with the highest validation IoU coefficient. As can be seen in the figure, training, and validation IoU rates were 87.43% and 74.21% respectively. As it was mentioned, in the training the original ZACAF model was used as pre-trained weights. In figure 6 on the right the performance of the model without transfer learning can be seen. As it can be inferred from the figure the transfer learning improved validation IoU metric by 25.45% which is significant.

3.9. Assessment of the performance of the framework for EF in the test set

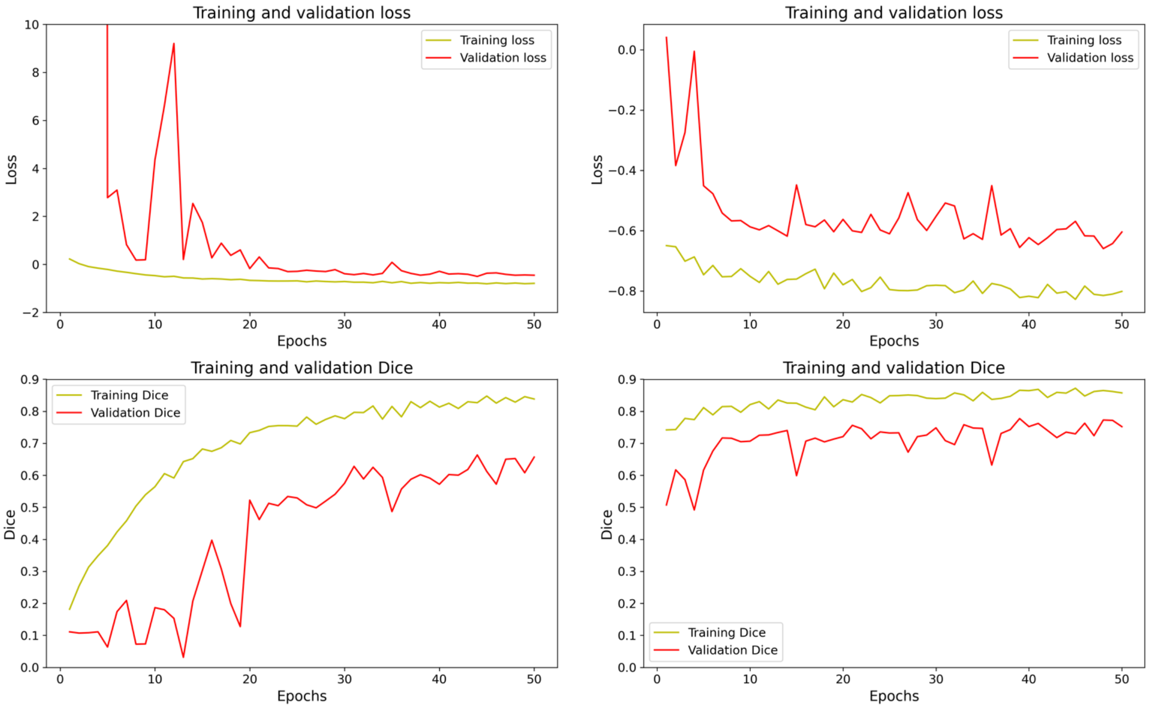

Figure 10: Evaluation of the performance of the model with(left) and without(right) using transfer learning. As it can be seen in the plots, training without transfer learning is far less accurate than the model that used the original ZACAF model as pre-trained weights and was then trained on the new dataset with Nrap videos.

To evaluate the framework's effectiveness in calculating EF, we compared the results obtained from manual assessment by an expert with those generated by the software. As EF calculation requires finding the area in all frames of a video to determine the ED and ES areas, the framework's performance was assessed using a series of frames from a test video, rather than random images from a validation set. To this end, we evaluated framework's performance with EF calculation, using 4 wildtype zebrafish embryos and 4 Nrap fish as inputs for the test set, without using them in the training. We first performed manual processing and estimation for each video to derive EF as the ground truth. Then, the model predicted ventricle masks for each frame of the input video, and the frames with the maximum and minimum area of the segmented ventricle were identified as the ES and ED frames, respectively. The framework then calculated EF and saved it, along with other indices, in a CSV file. The average absolute errors and standard deviations for the calculated EF of the 8 test videos, compared to the expert's manual calculation, were 7.23% and 4.78%, respectively. Additionally, TTA was applied to the model which resulted in a 3.21% improvement in the error rate. In conclusion, the TTA and transfer learning improved the model significantly and the final average error rate for the test set was 4.02%.

Here in figure 7 an example image from the validation set that was applied to the TTA can be seen.

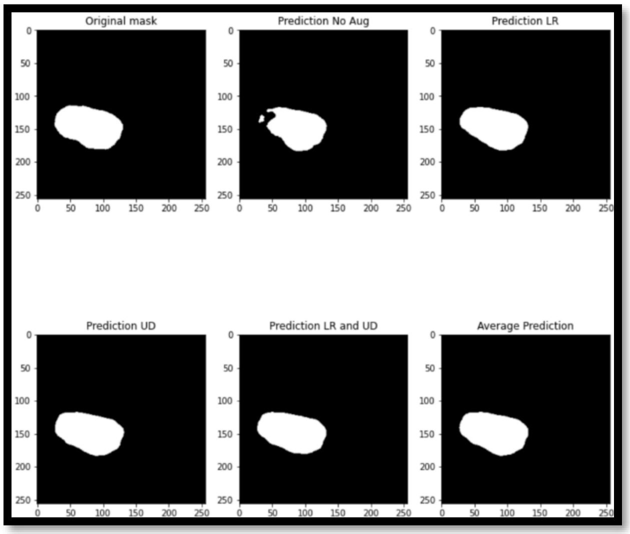

Figure 11: In this figure original masks and predictions resulted from various TTA transforms can be observed. On the top from left to right we can see, original mask created manually, prediction of the model on an image with no augmentation, prediction of the model on an image with left to right flip augmentation. And on the bottom from left to right you can see prediction of the model on an image with up to down flip augmentation, prediction of the model on an image with left to right flip and up to down augmentation, and the average of all predictions. As can be seen, the average prediction improved the initial prediction significantly using TTA. Here the IoU metric was improved from 74% to 93% just by using TTA.

Chapter 4: Discussions and Conclusion

4.1. Estimations in 2-D videos

The approach with microscopic videos relies on 2D videos to derive 3D volume estimation assuming the ventricle as a perfect ellipsoid. This assumption will result in accuracies in the measurements. Especially in mutant type where the shape of the chambers is not close to being ellipses this is going to be influential. However, the only solution to this problem is 3D imaging. In literature there are several studies that have extensively used 3D imaging technics like Z-stack imaging. In segmentation of the chambers using deep learning the third dimension is just going to be an extra layer of input. Granted the model is going to be more complicated but the concept is the same.

4.2. Consistency of measurement

Looking at the two frameworks that used deep learning for automatic segmentation of the zf heart, this method shows to be promising. The fully automated frameworks do the manual quantification of the cardiovascular metrics in a fraction of the time that it takes to do it manually. Additionally, manual segmentation is not consistent. Segmentation of the ventricle in these videos is a challenging task, even manually. The small size, ambiguous edges, and partial obstruction of the heart in the videos can also add complications to manual detection. We have investigated this quantitatively. We asked two experts to segment and measure the ventricle area in single frames of 12 sample videos. They were instructed to do the measurement twice for each frame manually with a short break between each try. The results were 12 frames, each measured 4 times. The standard deviation for each frame measurement was calculated, and the average of standard deviations of the measurements in these 12 frames was about 150 pixels with 50 pixels standard deviation. This is approximately 8% of an average size ventricular area in our setting’s scale. This shows the inconsistency in the manual segmentation. This could be especially significant with mutant embryos whose EF is usually very small. However, due to the nature of neural networks, trained models like ZACAF are consistent, which means that the measurement of a frame multiple times will always result in only one consistent measurement.

4.3. Frame rate issue

It is noteworthy to mention since the ground truth is created using the same frames for segmentation of the ventricle, the frame rate isn’t assessable. The ES and ED frames are the most important frames when it comes to the quantification of parameters like HR, EF, and FS. While recording the videos, the camera shutter takes a sequence of images with a certain fps. The higher the video's fps, the higher chance for exact ES and ED stages being recorded. This fact cannot be proved using the metrics because the prediction is only being compared with the existing manually segmented ground truth, and if the low fps causes the loss of ED or ES frames, there is no way to show it with the metrics.

4.4. Mutant fish lines challenges

From the segmentation point of view, there are two significant differences between the mutant and wildtype fish. The ventricle and the heart, in general, have abnormal shapes in several mutant types. In TTNtv case here, EF is much lower in the TTNtv model as the shape as well as the contractility are significantly affected. Thus, the ventricle area difference in ES and ED frames in TTNtv mutants is very low. **Figure 4** provides examples to compare wild and TTNtv zebrafish. In some cases, the ventricle is barely beating so that the area difference in ED and ES frames is lower than the segmentation error.

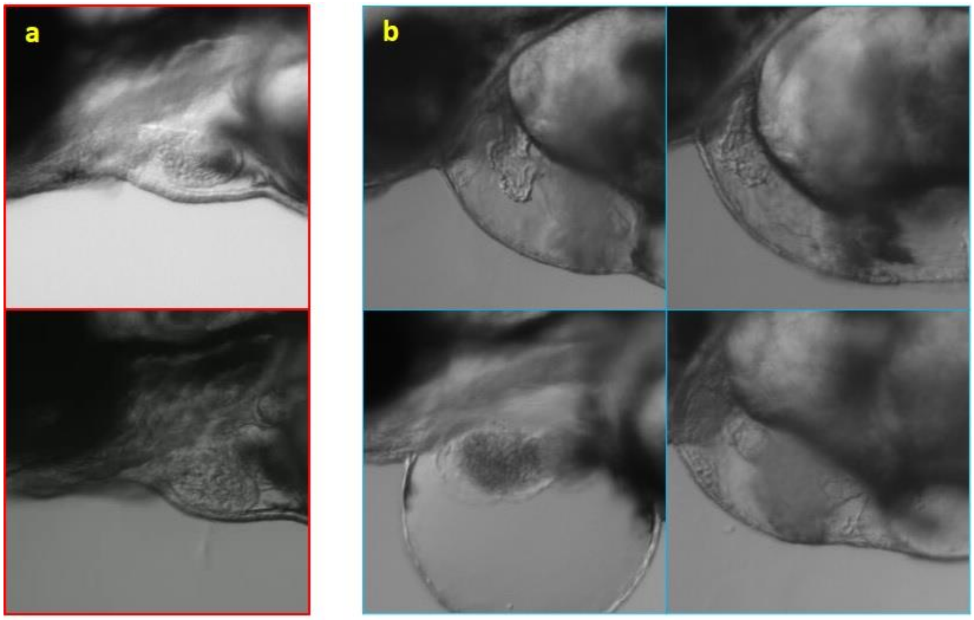

*Figure 12: **Comparison of the shape and size of wildtype (a) and TTNtv mutant zebrafish (b)**. Besides the abnormal shape of the heart with the swollen ventricular wall, the smaller size of the ventricle is also found with TTNtv mutants. Further, the swollen chest can be also noticed.*

In other words, the ventricle area hardly changes to the point that occasionally, the nominator of the formula of EF is lower than the estimation error. That is the primary source for the inaccuracies with the TTNtv mutants, and further improvements of preprocessing or optimization of the framework will not affect the result with the mutant significantly. The videos used in this work have low resolution in order to demonstrate the capability of our framework. Although this is beneficial for researchers to reduce required storage capacity, higher resolution would help resolve this issue, thus improving the robustness and accuracy for TTNtv mutant and wildtype fish in general.

Here, our framework can help researchers quantify the cardiac functions and parameters of studied zebrafish with minimum manual engineering efforts. In EF derivation, counting the pixels is more relevant and accurate than finding the long axis, which can be complicated since the ventricle is not a perfect ellipse. Further, the tool that most researchers use in the

ImageJ software is a freehand ruler, which could introduce inaccuracy, especially with the small size of heart chambers.

4.5. Pear shape of the ventricle in some videos

The pear-shaped ventricles showed no significant correlation to the genotypes. We speculate that this shape is observed due to the improper placement of the fish under the microscope. Hence it might be caused by the fish being placed not perfectly on its side. Considering that the ventricle is not a perfect ellipsoid, this shape could be the result of the perspective of the camera from the ventricle which can be imaged as the 2D shape similar to a pear instead of being closer to an ellipse.

4.6. Comparison between EF formulas

In formula (3), A is the 2D area calculated directly from the segmented ventricle and D_L is the long axis. This way of calculation of the volume does not assumes the shape of the ventricle to be a prolate spheroidal unlike formula (2). This formula is useful specifically for mutant fish where the long and short axis might not change significantly however the abnormal shape of the heart might contribute to an abnormal EF measurement. The results were calculated using both formulas. The average difference between the EF measurements using the two formulas was 3.34% which is negligible. However, some videos showed significant differences of up to 19.5%.

4.7. Discussion on transfer learning

The use of transfer learning in the ZACAF model has proved to be a successful technique for improving the model's performance for ventricle segmentation in zebrafish. By utilizing pre-trained weights from the original model, the new model was able to benefit from the features learned during the previous training, resulting in faster training and better performance than training the model from scratch on the new dataset. Additionally, the use of callbacks such as the model checkpoint helped to ensure that the best-performing model was saved and used for further analysis, allowing for the model to continue improving its performance.

4.8. Discussion on TTA

In this case, TTA was beneficial because it increased the variability of the test data and helped to reduce the effect of any biases in the original dataset. Since the ZACAF model was trained on a different dataset and microscope setup, there could be some differences in the characteristics of the new dataset that were not present in the original dataset. By applying TTA, model was able to generate additional test data that had different characteristics, which helped to reduce the impact of any biases in the original dataset. Another benefit of TTA is that it can help to increase the robustness of the model to variations in the input data. Since the model is exposed to a larger variety of test data during inference, it is less likely to overfit to a particular type of input and more likely to generalize well to new data. However, TTA also has some potential drawbacks. One of the main drawbacks is that it can increase the computational cost of making predictions since the model needs to process multiple versions of the test data. Depending on the complexity of the model and the number of test data versions generated, the computational cost can be significant. Another potential drawback of TTA is that it can introduce some variability into the predictions, which can make it difficult to interpret the results. Since the final prediction is based on an average of multiple predictions, it may not be clear which version of the test data was responsible for a particular prediction. This can make it challenging to identify specific areas of the input data that the model is struggling with.

4.9. Conclusion

Based on the findings presented in this paper, downregulation of Nebulin Related Anchoring Protein (NRAP) in embryonic zebrafish does not significantly affect cardiac function, specifically ejection fraction and fractional shortening, as measured by the Zebrafish Automatic Cardiovascular Assessment Framework (ZACAF) based on a U-net deep learning model. While NRAP has been shown to play a role in myofibril assembly in cardiac and skeletal muscle, and overexpression of NRAP has been associated with severe skeletal muscle myopathy in zebrafish, our results do not support the hypothesis that NRAP downregulation would result in improved cardiac function.

The modified ZACAF deep learning model demonstrated effectiveness in implementing data from new research teams with different recording setups and phenotypes and provided a fully automated and accurate measurement of cardiovascular parameters, which could be useful in future research. The effectiveness of Transfer learning, Data augmentation, Test time augmentation were evaluated. Considerations for video recording and preprocessing, zebrafish orientation and handling, and deep learning model architecture were discussed.

Overall, while previous studies have suggested the potential therapeutic advantages of NRAP downregulation in multiple model organisms, our findings do not support this hypothesis in embryonic zebrafish. Further research is needed to understand the role of NRAP in cardiac function and its potential as a therapeutic target for cardiomyopathy.

References

1. Wisneski, J., et al., *Left ventricular ejection fraction calculated from volumes and areas: underestimation by area method.* Circulation, 1981. **63**(1): p. 149-151.
2. Ling, D., et al., *Quantitative measurements of zebrafish heartrate and heart rate variability: A survey between 1990–-2020.* Computers in Biology and Medicine, 2021: p. 105045.
3. Maragos, P., *Chapter 13 - Morphological Filtering*, in *The Essential Guide to Image Processing*, A. Bovik, Editor. 2009, Academic Press: Boston. p. 293-321.
4. Pizer, S.M., et al., *Adaptive histogram equalization and its variations.* Computer Vision, Graphics, and Image Processing, 1987. **39**(3): p. 355-368.
5. Canny, J., *A Computational Approach to Edge Detection.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986. **PAMI-8**(6): p. 679-698.
6. Akerberg, A.A., et al., *Deep learning enables automated volumetric assessments of cardiac function in zebrafish.* Disease models & mechanisms, 2019. **12**(10): p. dmm040188.
7. Dhanachandra, N., K. Manglem, and Y.J. Chanu, *Image Segmentation Using K -means Clustering Algorithm and Subtractive Clustering Algorithm.* Procedia Computer Science, 2015. **54**: p. 764-771.
8. Bishop, C.M. and N.M. Nasrabadi, *Pattern recognition and machine learning*. Vol. 4. 2006: Springer.
9. Gupta, L. and T. Sortrakul, *A gaussian-mixture-based image segmentation algorithm.* Pattern Recognition, 1998. **31**(3): p. 315-325.
10. Wessells, R.J. and R. Bodmer, *Screening assays for heart function mutants in Drosophila.* Biotechniques, 2004. **37**(1): p. 58-66.
11. Nasrat, S., et al., *Semi-automated detection of fractional shortening in zebrafish embryo heart videos.* Current Directions in Biomedical Engineering, 2016. **2**(1): p. 233-236.
12. Huang, W.-Y., et al., *Transgenic expression of green fluorescence protein can cause dilated cardiomyopathy.* Nature medicine, 2000. **6**(5): p. 482-483.
13. Naderi, A.M., et al., *Deep learning-based framework for cardiac function assessment in embryonic zebrafish from heart beating videos.* Computers in biology and medicine, 2021. **135**: p. 104565.
14. Ronneberger, O., P. Fischer, and T. Brox. *U-net: Convolutional networks for biomedical image segmentation*. in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. 2015. Springer.
15. Decourt, C. and L. Duong, *Semi-supervised generative adversarial networks for the segmentation of the left ventricle in pediatric MRI.* Computers in Biology and Medicine, 2020. **123**: p. 103884.
16. Jadon, S. *A survey of loss functions for semantic segmentation*. in *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*. 2020.
17. Merlo, M., et al., *Evolving concepts in dilated cardiomyopathy.* Eur J Heart Fail, 2018. **20**(2): p. 228-239.
18. Hershberger, R.E., D.J. Hedges, and A. Morales, *Dilated cardiomyopathy: the complexity of a diverse genetic architecture.* Nat Rev Cardiol, 2013. **10**(9): p. 531-47.
19. Wheeler, F.C., et al., *QTL mapping in a mouse model of cardiomyopathy reveals an ancestral modifier allele affecting heart function and survival.* Mamm Genome, 2005. **16**(6): p. 414-23.
20. Hoage, T., Y. Ding, and X. Xu, *Quantifying cardiac functions in embryonic and adult zebrafish.* Methods Mol Biol, 2012. **843**: p. 11-20.

21. Bland, J.M. and D.G. Altman, *Statistical methods for assessing agreement between two methods of clinical measurement.* Lancet, 1986. **1**(8476): p. 307-10.
22. Lu, S., D.E. Borst, and R. Horowits, *Expression and alternative splicing of N-RAP during mouse skeletal muscle development.* Cell Motil Cytoskeleton, 2008. **65**(12): p. 945-54.
23. Jirka, C., et al., *Dysregulation of NRAP degradation by KLHL41 contributes to pathophysiology in nemaline myopathy.* Hum Mol Genet, 2019. **28**(15): p. 2549-2560.
24. Lu, S., et al., *Cardiac-specific NRAP overexpression causes right ventricular dysfunction in mice.* Exp Cell Res, 2011. **317**(8): p. 1226-37.
25. Truszkowska, G.T., et al., *Homozygous truncating mutation in NRAP gene identified by whole exome sequencing in a patient with dilated cardiomyopathy.* Sci Rep, 2017. **7**(1): p. 3362.
26. Shorten, C. and T.M. Khoshgoftaar, *A survey on Image Data Augmentation for Deep Learning.* Journal of Big Data, 2019. **6**(1): p. 60.
27. Cossio, M., *Augmenting Medical Imaging: A Comprehensive Catalogue of 65 Techniques for Enhanced Data Analysis.* arXiv preprint arXiv:2303.01178, 2023.
28. Pan, S.J. and Q. Yang, *A Survey on Transfer Learning.* IEEE Transactions on Knowledge and Data Engineering, 2010. **22**(10): p. 1345-1359.
29. Moshkov, N., et al., *Test-time augmentation for deep learning-based cell segmentation on microscopy images.* Scientific Reports, 2020. **10**(1): p. 5068.