



Article

# Image Segmentation for Mitral Regurgitation with Convolutional Neural Network Based on UNet, Resnet, Vnet, FractalNet and SegNet: A Preliminary Study

Linda Atika<sup>1,2</sup> , Siti Nurmaini<sup>3,\*</sup> , Radiyati Umi Partan<sup>4</sup> and Erwin Sukandi<sup>5</sup>

<sup>1</sup> Doctoral Program of Engineering Science, Faculty of Engineering, Universitas Sriwijaya, Palembang 30128, Indonesia

<sup>2</sup> Department of Computer Science, Universitas Bina Darma, Palembang 30264, Indonesia

<sup>3</sup> Intelligent System Research Group, Universitas Sriwijaya, Palembang 30128, Indonesia

<sup>4</sup> Internal Medicine Departement, Faculty of Medicine, Universitas Sriwijaya, Palembang 30128, Indonesia

<sup>5</sup> Cardiology Division, Internal Medicine Departement, Faculty of Medicine, Dr. Mohammad Hoesin Hospital, Universitas Sriwijaya, Palembang 30128, Indonesia

\* Correspondence: siti\_nurmaini@unsri.ac.id

**Abstract:** The heart's mitral valve is the valve that separates the chambers of the heart between the left atrium and left ventricle. Heart valve disease is a fairly common heart disease, and one type of heart valve disease is mitral regurgitation, which is an abnormality of the mitral valve on the left side of the heart that causes an inability of the mitral valve to close properly. Convolutional Neural Network (CNN) is a type of deep learning that is suitable for use in image analysis. Segmentation is widely used in analyzing medical images because it can divide images into simpler ones to facilitate the analysis process by separating objects that are not analyzed into backgrounds and objects to be analyzed into foregrounds. This study builds a dataset from the data of patients with mitral regurgitation and patients who have normal hearts, and heart valve image analysis is done by segmenting the images of their mitral heart valves. Several types of CNN architecture were applied in this research, including U-Net, SegNet, V-Net, FractalNet, and ResNet architectures. The experimental results show that the best architecture is U-Net3 in terms of Pixel Accuracy (97.59%), Intersection over Union (86.98%), Mean Accuracy (93.46%), Precision (85.60%), Recall (88.39%), and Dice Coefficient (86.58%).

**Keywords:** Convolutional Neural Network (CNN); mitral regurgitation; segmentation; unet



**Citation:** Atika, L.; Nurmaini, S.; Partan, R.U.; Sukandi, E. Image Segmentation for Mitral Regurgitation with Convolutional Neural Network Based on UNet, Resnet, Vnet, FractalNet and SegNet: A Preliminary Study. *Big Data Cogn. Comput.* **2022**, *6*, 141. <https://doi.org/10.3390/bdcc6040141>

Academic Editors: Dilbag Singh, Heung-No Lee and Vijay Kumar

Received: 31 October 2022

Accepted: 23 November 2022

Published: 25 November 2022

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



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Every year in the United States, the population experiencing heart attacks or strokes is recorded at more than 2 million people. The biggest cause of death is cardiovascular disease [1]. Each year, the Centers for Disease Control and Prevention, the National Institutes of Health in collaboration with the American Heart Association (AHA), and other government agencies compile the latest statistics relating to heart disease, stroke, and cardiovascular and other metabolic diseases as outlined in the Heart Disease and Stroke statistics update. From data collected by Coronary Artery Risk Development in Young Adults (CARDIA), Atherosclerosis Risk in Communities (ARIC), and Cardiovascular Health Study (CHS) in 2000 in the United States in the adult population; the incidence rate of heart valve disease is the most common is mitral regurgitation, which is 1.7%, and increased from 0.7% in participants aged 18 to 44 years to 11.3% in participants aged over 75 years. The incidence of regurgitant mitral valve disease is estimated to be four times higher than that of stenotic aortic valve disease [2].

The heart has four heart valves, which are responsible for circulating blood and producing a pulse. Examination of the presence or absence of heart disease can be done using an echocardiography tool that can display the four heart chambers and heart valves.

Echocardiogram examination in the presence of color Doppler is a fairly robust and convincing imaging method to evaluate the geometry, dynamics and function of degenerative and functional mitral valve (MV) regurgitation. In addition, the presence of color Doppler is useful for medical personnel to identify the location of the regurgitation hole and the severity of mitral regurgitation. Automated assessment of mitral regurgitation using color doppler echocardiographic images to assess the severity of the mitral disease is of great value in helping surgeons perform MV repair [3,4].

The most common congenital heart disease found is Atrial Septal Defect (ASD). The relationship between ASD and valvular heart disease has been recognized for years. In Indonesia, a study conducted at Dr. Sardjito Hospital showed records from echocardiography examination, the echocardiograph was Vivid 7, with a total of 103 adult patients, consisting of 16 men and 87 women aged between 17 to 76 years, with an average age of 36 years, found to have ASD [5].

Deep learning is one type of machine learning that has become quite popular in recent years. Deep learning is also suitable for research in the health sector because it can process large amounts of data and can accept several types of data, such as images containing a collection of input points, then produce the appropriate output. The research of Andre Esteva et al. presents learning techniques in computer vision for health care that impact several health fields, especially medical imaging. This study outlines how deep learning models can be used on large data sets because their capabilities run on dedicated computing hardware and can accept multiple data types as suitable inputs in the health sector, because health data has heterogeneous input data [6].

Segmentation, feature extraction, and classification tend to use the advantages of Artificial Intelligence (AI), which uses neural network and deep learning techniques to get more accurate results in segmentation, feature extraction, and classification in the fields of medical diagnostics [7]. In recent years, many studies have used deep learning approaches, especially Convolutional Neural Networks, for the detection of diseases. Another study also used a deep learning approach for the automatic detection of melanoma skin cancer from dermoscopic skin samples, which accurately classifies malignant vs. benign melanoma. This study used dermoscopic images containing different cancer samples. The data obtained were from the International Skin Imaging Collaboration data repository (ISIC 2016, ISIC2017, and ISIC 2020 Evaluation of models based on accuracy, precision, recall, specificity, and F1 score. The Deep Convolutional Neural Network (DCNN) classifier achieves high accuracy [7].

One way to diagnose the disease is segmentation. Segmentation is one of the keys to conducting medical image analysis. Segmentation is the process of partitioning a digital image in some area, which means simplifying or transforming the representation of an image into something more meaningful, that is easier to analyze and identify. Many algorithms have been used for image segmentation, U-Net is an image segmentation technique developed especially for image segmentation tasks. In the medical imaging community, U-Net has been accepted as the main tool for segmentation tasks in medical imaging.

The success of U-Net has been proven by its use with such images as those from CT scans, MRIs, x-rays, and microscopy [8]. Meanwhile, the study of B. Ait Skourt et al. proposed the segmentation of CT images of the lungs with U-Net architecture. U-Net architecture is one of the architectures that are widely used in deep learning for image segmentation. The experimental results show that the segmentation architecture is accurate and the U-Net architecture provides an accurate picture of the lung in detecting lung cancer [9].

In the study developed by Ronneberger et al. [10], they proposed a network and training strategy with the efficient use of data augmentation on large annotated samples. The proposed method is the U-Net architecture that is being developed. In the proposed method, the structure of the method is seen using the encoder on the left and then performed using the decoder on the right. The encoder is one of the characteristics of the CNN structure that is used to compose the Convolution layer [10]. Convolutional Neural Network has

been applied in various medical image segmentation tasks and has been shown to perform better than traditional algorithms [11]. In the research conducted by Q. Zhang et al., using the merging method based on the Resnet and U-Net architectures on Ultrasound Nerve data using Residual Units on the U-net architecture, the combination of these methods got the best results at the Dice Coefficient (69.15%) [12]. In this paper, we also used U-Net development, which is U-Net3, and we obtained 86.58% higher Dice Coefficient results compared to that study.

In the research developed by Daniel et al., the proposed design of the new Unet architecture, namely Unet3, was introduced to detect and track people in crowded and continuous environments, such as airports or stations. In this architecture, the addition of normalized batches after the activation of the first relay and after the max-pooling and upsampling functions was proposed. This approach was tested on public datasets, TVHeads Dataset, which resulted in an F1 Score output ranging from 90%. They also performed performance comparisons between U-Net3 with FractalNet, Resnet, U-Net, U-Net2, and Segnet. The results show that U-Net3 is superior to other architectures [13]. Referring to the results of this research, we built the proposed model, which is U-Net3, but the difference is that we used the medical data that we built ourselves, not the public data. We also added one other architecture, which is Vnet, to compare our proposed model.

Another study conducted by Numaini et al. proposed a CNN-based U-Net architecture to automatically segment cardiac chambers to detect abnormalities (holes) in the cardiac septum by using a segmentation model for four classes, by comparing the performance of two architectures, namely U-Net and V-net. The results showed that the accuracy of the two architectures was above 90%, while U-Net had better accuracy than the V-Net model architecture; therefore, it can be concluded that the CNN architecture succeeded in segmenting the heart chambers for the detection of defects in the cardiac septum and can support the work of heart experts [14]. In this study, we compared the proposed model and one other architectural model, while we compared the proposed model with six other architectural models.

Other medical research conducted by Kalane et al. used the U-Net architecture to automatically detect COVID-19 disease. The dataset used uses 1000 chest CT images consisting of 448 images of patients with COVID-19 disease and 552 patients without COVID-19 disease. The dataset was obtained from the GitHub repository and the Italian Society of Excellent Collections of Medical Radiology and Interventions. The experimental results show that the U-Net architecture is proven to be effective and produces output with an overall accuracy of 94.10% [15]. That study discusses the COVID-19 disease, as we discuss Mitral Regurgitation disease.

Based on the description above, in this paper, we also propose a CNN-based architecture, using a segmentation model on a four-chamber heart image, to make a segmentation model using valvular heart disease and mitral regurgitation image datasets. There are many articles on medical image segmentation, especially heart disease, in various fields of image research, but very little research on segmentation on heart valve disease. Innovations or updates from this paper include:

In the research developed by Daniel et al., the proposed design of the new Unet is building a new dataset, namely Mitral Regurgitation, consisting of:

- Designing a CNN model for segmentation of Mitral Regurgitation heart valve disease with high Accuracy
- Developed CNN-based U-Net3 architecture for segmentation of Mitral Valve disease and normal valve condition
- Validating U-Net3 model with six other architectures using pixel accuracy, intersection over union, mean accuracy, precision, recall, and dice coefficient

Section 2 of this paper consists of patients with mitral regurgitation disease and normal patients. The focus of this paper is to make a segmentation from proposed datasets using labeled images. In Table 1 we describe details amount of filtered frame from each video.

**Table 1.** No. of images for training and testing.

Video	No. of Videos	Total Duration (second)	Frame Rate (fps)	Filter Frames
Normal 1	1 Video	02: 00	25	28
Normal 2	1 Video	02: 00	25	28
Normal 3	2 Videos	04: 00	25	41
Normal 4	2 Videos	04: 00	25	38
Normal 5	1 Video	02: 00	25	22
Normal 6	1 Video	02: 00	25	12
Normal 7	1 Video	02: 00	25	7
Normal 8	1 Video	02: 00	25	13
Normal 9	1 Video	02: 00	25	25
Normal 10	1 Video	02: 00	25	15
Normal 11	1 Video	02: 00	25	11
Normal 12	2 Videos	04: 00	25	41
Normal 13	1 Video	02: 00	25	13
Normal 14	1 Video	02: 00	25	11
Normal 15	1 Video	02: 00	25	17
Normal 16	2 Videos	02: 00	25	41
Normal 17	1 Video	02: 00	25	16
Normal 18	1 Video	02: 00	25	7
Normal 19	1 Video	02: 00	25	30
Normal 20	1 Video	02: 00	25	22
Normal 21	1 Video	02: 00	25	31
Mitral Regurgitation 1	1 Video	02: 00	25	16
Mitral Regurgitation 2	1 Video	02: 00	25	6
Mitral Regurgitation 3	1 Video	02: 00	25	5
Mitral Regurgitation 4	1 Video	02: 00	25	11
Mitral Regurgitation 5	1 Video	02: 00	25	22
Mitral Regurgitation 6	1 Video	02: 00	25	4
Mitral Regurgitation 7	1 Video	02: 00	25	28
Mitral Regurgitation 8	1 Video	02: 00	25	18
Mitral Regurgitation 9	3 videos	06: 00	25	52
Mitral Regurgitation 10	1 Video	02: 00	25	12
Mitral Regurgitation 11	1 Video	02: 00	25	23
Mitral Regurgitation 12	1 Video	02: 00	25	30
Mitral Regurgitation 13	1 Video	02: 00	25	41
Mitral Regurgitation 14	1 Video	02: 00	25	31
Mitral Regurgitation 15	1 Video	02: 00	25	36
Mitral Regurgitation 16	1 Video	02: 00	25	14
Mitral Regurgitation 17	1 Video	02: 00	25	7
Mitral Regurgitation 18	1 Video	02: 00	25	32
Mitral Regurgitation 19	1 Video	02: 00	25	26
Mitral Regurgitation 20	1 Video	02: 00	25	21
Mitral Regurgitation 21	1 Video	02: 00	25	19
Total Frames	923			

In Table 2 we describe the details amount of data for training and testing, and unseen data.

**Table 2.** Dataset for training, testing, and unseen.

No	No. of Patients	Total Frames	Training	Testing	Unseen
Normal	21	469	334	79	56
Mitral Regurgitation	21	454	287	77	90
Total	42	923	621	156	146

In Table 3, we also describe parameters for the various CNN Architecture.

In Table 4 we describe details of the architecture of U-Net.

**Table 3.** Parameters for the various CNN Architectures.

Architecture	Batch Size	Learning Rate	Epoch
SegNet	64	0.00001	500
ResNet	64	0.00001	500
U-Net	64	0.00001	500
U-Net 2	64	0.00001	500
U-Net 3	64	0.00001	500
V-Net	64	0.00001	500
Fractal-Net	64	0.00001	500

**Table 4.** U-Net Detail Architecture.

Layer	Kernel Size	Stride	Activation Function	Output
Input Layer	-	-	-	$256 \times 256 \times 1$
Convolution Layer 1	$64 \times 64 \times 1$	1	ReLU	$128 \times 128 \times 3$
Batch Normalization				
Convolution Layer 2	$64 \times 64 \times 1$	1	ReLU	$128 \times 128 \times 3$
Max Pooling 1	$2 \times 2$	2		$128 \times 128 \times 3$
Batch Normalization				
Convolution Layer 3	$128 \times 128 \times 3$	1	ReLU	$256 \times 256 \times 3$
Batch Normalization				
Convolution Layer 4	$128 \times 128 \times 3$	1	ReLU	$256 \times 256 \times 3$
Max Pooling 2	$2 \times 2$	2		$256 \times 256 \times 3$
Batch Normalization				
Convolution Layer 5	$256 \times 256 \times 3$	1	ReLU	$512 \times 512 \times 3$
Batch Normalization				
Convolution Layer 6	$256 \times 256 \times 3$	1	ReLU	$512 \times 512 \times 3$
Max Pooling 3	$2 \times 2$	2	-	$512 \times 512 \times 3$
Batch Normalization				
Convolution Layer 7	$512 \times 512 \times 3$	1	ReLU	$1024 \times 1024 \times 3$
Batch Normalization				
Convolution Layer 8	$512 \times 512 \times 3$	1	ReLU	$1024 \times 1024 \times 3$
Max Pooling 4	$2 \times 2$	2	-	$1024 \times 1024 \times 3$
Batch Normalization				
Convolution Layer 9	$1024 \times 1024 \times 3$	1	ReLU	$512 \times 512 \times 3$
Batch Normalization				
Convolution Layer 10	$1024 \times 1024 \times 3$	1	ReLU	$512 \times 512 \times 3$
Up 1	$512 \times 512 \times 3$	3 (axis)	ReLU	$512 \times 512 \times 3$
Batch Normalization				
Convolution Layer 11	$512 \times 512 \times 3$	1	ReLU	$256 \times 256 \times 3$
Batch Normalization				
Covolutional Layer 12	$512 \times 512 \times 3$	1	ReLU	$256 \times 256 \times 3$
Up 2	$256 \times 256 \times 3$	3 (axis)	ReLU	$256 \times 256 \times 3$
Batch Normalization				
Covolutional Layer 13	$256 \times 256 \times 3$	1	ReLU	$128 \times 128 \times 3$
Batch Normalization				
Covolutional Layer 14	$256 \times 256 \times 3$	1	ReLU	$128 \times 128 \times 3$
Up 3	$128 \times 128 \times 3$	3 (axis)	ReLU	$128 \times 128 \times 3$
Batch Normalization				
Convolutional Layer 15	$128 \times 128 \times 3$	1	ReLU	$64 \times 64 \times 3$
Batch Normalization				
Covolutional Layer 16	$128 \times 128 \times 3$	1	ReLU	$64 \times 64 \times 3$
Up 4	$64 \times 64 \times 3$	1	ReLU	$64 \times 64 \times 3$
Batch Normalization				
Covolutional Layer 17	$64 \times 64 \times 3$	1	ReLU	
Convolution Layer 18	$64 \times 64 \times 3$	1	Hard_Sigmoid	$2 \times 2 \times 3$
Batch Normalization				
Output Layer	-	-	Hard_Sigmoid	1

The performance assessment was conducted on pixel accuracy, intersection over union, and dice coefficient values and also compared the proposed architecture, U-Net3, with six other architectures, namely SegNet, Resnet, U-Net, U-Net2, and V-Net, which can be seen in Table 5.

**Table 5.** Performance measurement with splitting data.

Evaluation Metrics	Performance (%)						
	Segnet	ResNet	U-Net	U-Net2	U-Net3	FractalNet	VNet
Pixel Accuracy	97.62	97.40	97.58	96.09	97.59	97.63	97.67
IoU	87.20	86.04	86.78	78.50	86.98	87.03	87.23
Mean Accuracy	93.73	92.73	92.59	83.73	93.46	92.91	93.05
Precision	85.49	84.69	86.92	85.76	85.60	85.49	87.01
Recall	88.95	86.99	86.46	68.57	88.39	88.95	87.38
Dice Coefficient	86.85	85.48	86.34	75.36	86.58	86.85	86.87

In Table 6 showing performance measurement with unseen data. From the results of this study, it can be concluded that the U-Net3 architecture has a fairly effective and best performance among other comparison architectures.

**Table 6.** Performance measurement with unseen data.

Evaluation Metrics	Performance (%)						
	Segnet	ResNet	U-Net	U-Net2	U-Net3	FractalNet	VNet
Pixel Accuracy	97.05	96.39	96.79	95.69	97.24	96.61	94.53
IoU	85.26	82.58	84.45	78.33	86.44	82.97	74.18
Mean Accuracy	88.69	88.91	90.87	83.12	86.92	87.90	81.61
Precision	87.24	84.16	84.82	87.37	80.44	88.64	74.40
Recall	82.62	79.46	83.38	67.29	86.16	76.89	65.52
Dice Coefficient	84.89	81.29	83.72	75.44	86.14	81.91	67.73

## 2. Materials and Methods

The Mitral Regurgitation Valve Segmentation process begins with the collection of a dataset of mitral regurgitation valve disease and normal heart valves, where the data collected are in the form of color Doppler echocardiogram videos obtained from patients suspected of having heart valve abnormalities. Furthermore, from the videos, data preparation is carried out by breaking the videos into a collection of images. The annotation process is carried out on each image to produce a ground truth. After all the images are annotated, the segmentation and prediction process is carried out.

### 2.1. Data Acquisition

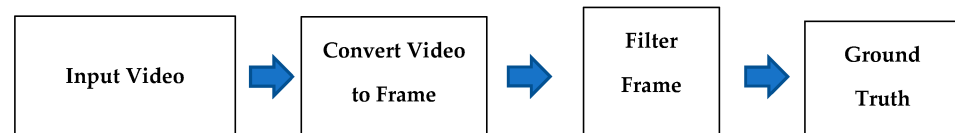
In this research, a private dataset was built. The dataset was built by taking video echocardiography recording data consisting of mitral valve heart disease, regurgitation, and normal hearts, in the form of a four-chamber display in video format in the form of \*.avi. Video data retrieval was taken from the Mohammad Hoesin Hospital in Palembang for the period of December 2019 to December 2021 using Transthoracic Echocardiogram (TTE) [2].

The data collected consisted of 42 patients and 923 images. We only used images of good quality and also frames that showed all parts of the heart. All 21 patients had mitral valve leakage, which resulted in 454 images, and 21 patients had normal hearts, which resulted in 469 images. The total number of frames obtained for the training and

testing data used was 777 images. Of the 621 images for training and 156 for testing, they were divided into 37 patients for training and testing, and 6 patients for unseen data. The new data from 6 different patients outside of training and testing produced 146 images. Four patients had mitral valve leakage, which resulted in 90 images, and two patients had normal hearts, which resulted in 56 images.

## 2.2. Data Pre-Processing

Data preprocessing is a series of steps in the data filtering process. The preprocessing process is shown in Figure 1.



**Figure 1.** Pre-Processing.

Mitral Regurgitation video data pre-processing for segmentation starts from converting videos in .avi format into a collection of images. The next step is to filter the data by taking data on images that have mitral regurgitation heart disease and normal heart when the heart valve is closed, the presence of color Doppler on echocardiogram examination will indicate the presence or absence of mitral valve disease. The next step in data pre-processing is to label the filtered frames. The labeling process or ground truth is carried out using the Label Me application to get ground truth results.

## 2.3. Model Architecture

The deep learning methods used in this research are the U-Net, Resnet, SegNet, and Vnet architectures, which are various types of CNN architectures designed to perform image processing. The U-Net architecture looks like a “U” with three parts: shrinkage, bottleneck, and expansion. The contraction section consists of many contraction blocks. Each block receives an input, applying two layers,  $3 \times 3$  convolution followed by  $2 \times 2$  maxpooling [2]. The ResNet architecture introduces a new block called a residual block. A residual Neural Network has the ability to pass through multiple layers using shortcuts, and can allow a layer to copy an input to the next layer. The ResNet architecture is a CNN architecture developed using a residual network. The main idea of the ResNet network is to introduce residual blocks. Residual blocks overlay constant mapping layers based on a flat grid to perform residual learning, improve feature extraction accuracy, and solve missing gradient problems [12]. V-Net is one of the most popular CNN architectures for medical imaging. V-Net is one of the architectures used to segment the image. The V-Net architecture has two important parts, namely the compression path and decompression path [16].

### 2.3.1. SegNet

SegNet has a corresponding encoder network and decoder network, followed by a final pixel classification layer. The network encoder consists of 13 correlated convolution layers responding to the first 13 convolution layers in the VGG16 network designed for object classification [17].

### 2.3.2. ResNet

ResNet is a CNN architecture designed with residual networks in mind. The main idea of the ResNet network is to introduce the remaining blocks. Residual blocks overlay constant mapping layers based on a flat grid to perform residual learning, improve feature extraction accuracy, and solve missing gradient problems [12].

### 2.3.3. V-Net

V-Net is one of the most popular CNN architectures for medical imaging. V-Net is one of the architectures used to segment the image. V-Net has two important parts of the architecture, namely compression and path decompression, a process that converts input data into an output data stream with a smaller size is called compression while the process of converting data that has been compressed into data originally is called decompression [16].

### 2.3.4. Fractal-Net

FractalNet is one of the architectures from CNN that avoids residual connections. This architecture involves the repeated application of simple expansion rules to create a fractal convolution network. The fractal network contains sub-paths that interact with different lengths. Each internal signal is altered by the filter before heading toward the next layer [18].

### 2.3.5. U-Net

U-Net is a CNN architecture developed for bio-medical image segmentation. This network is based on a fully connected network. The network consists of contract paths and extension paths, resulting in architecture in the form of U. U-Net consisting of convolution operations, maximum pooling, ReLU activation, upsampling layer, and downsampling. The downsampling path has five convolution blocks. Each block has a two-layer convolution with a  $3 \times 3$  filter [11].

### 2.3.6. U-Net3

U-Net3 is a U-Net architecture modified by adding a Batch Normalization function with ReLU activation, max pooling, and upsampling. U-Net3 was Introduced by Daniel Licioti in 2018. This architecture consists of two main parts, namely the contracting path on the left and the expansive path on the right. The first path corresponds to the convolution network architecture, which consists of looping applications of two  $3 \times 3$  convolutions, whereas the second path and each path are followed by ReLU and  $2 \times 2$  Maxpooling operations with stride 2 for downsampling. In each downsampling step, the feature is duplicated while in the expansive path, each upsampling step of the feature is followed by  $2 \times 2$  convolution and ReLU. At the end of the layer, it is followed by  $1 \times 1$  convolution, which is used to map each of the 32 feature vector components into a specified number of classes [13].

## 2.4. Performance Metric

The model produces image predictions, which are then evaluated. The grounded image data is used as a reference to measure the performance of the segmentation results. The performance of the segmentation results is measured by validating the ground truth image used as test data. This validation is carried out with *Pixel Accuracy* [19], *Intersection over Union* [20], *Mean Accuracy*, *Precision*, *Recall*, and *Dice Coefficient* [20] with Equations (1)–(6).

$$\text{Pixel Accuracy} = \frac{TP + TN}{TP + TN + FN + FP} \quad (1)$$

$$\text{IoU} = \frac{\text{Interseccion}}{\text{Union}} = \frac{TP}{TP + FN + FP} \quad (2)$$

$$\text{F1 Score/Dice Coefficient} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

$$\text{Mean Accuracy} = \frac{\sum \text{Pixel Accuracy of n objects}}{\text{n objects}} \quad (4)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$



$$Recall = \frac{TP}{TP + FN} \tag{6}$$

### 3. Results and Discussion

#### 3.1. Results

In this study, we demonstrate that a CNN-based U-Net architecture can successfully explain MR heart valve segmentation. This section will explain the results of the predictions on the testing data. Each CNN architecture is trained. After the training process is complete, the entire architecture is evaluated with different patient data that are not used in the training process. The performance of various architectures is analyzed with three parameters, namely Pixel Accuracy, Intersection over Union, and Dice Coefficient. Figure 2 shows the different between raw image and ground truth.

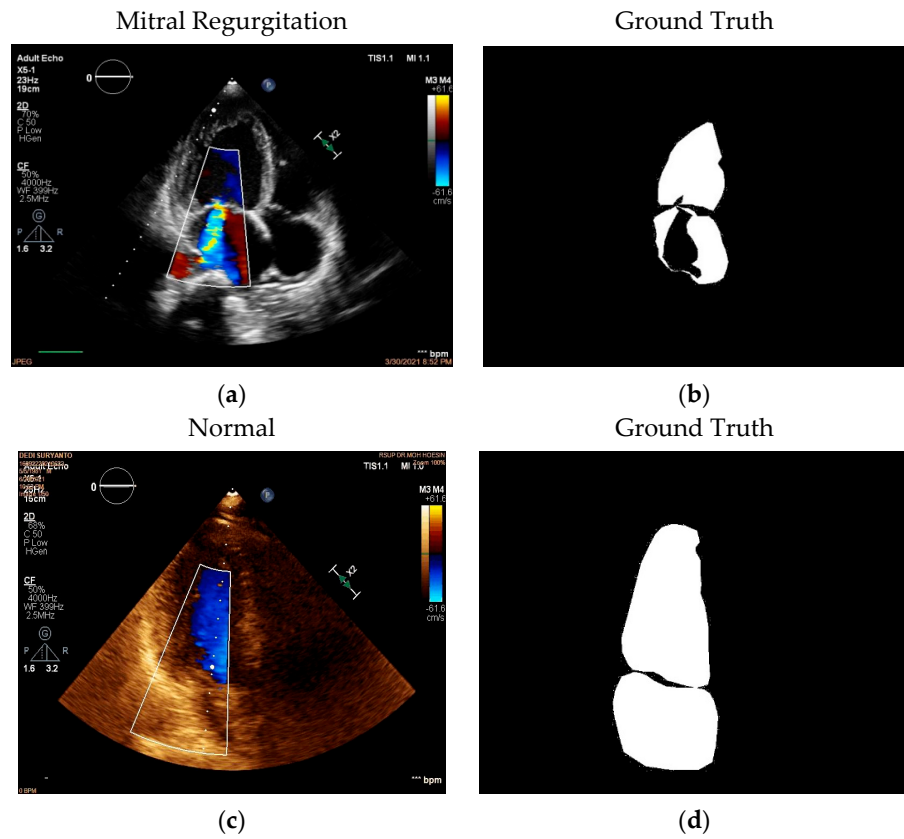


Figure 2. (a) Original image; (b) Ground Truth; (c) Normal; (d) Ground Truth.

Figure 3 shows the illustration of implemented U-Net3 architec.

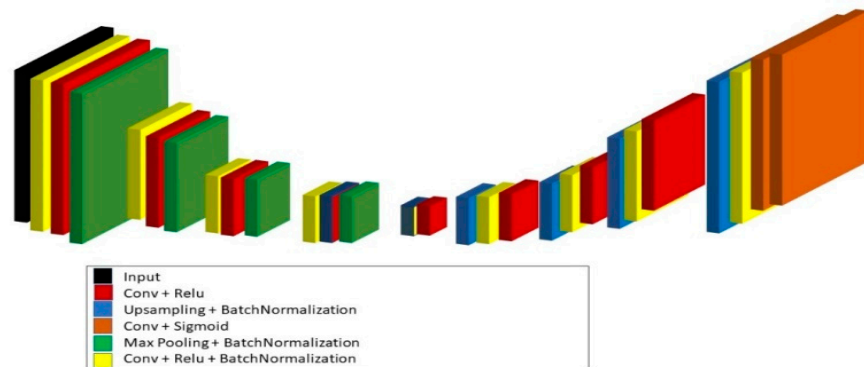
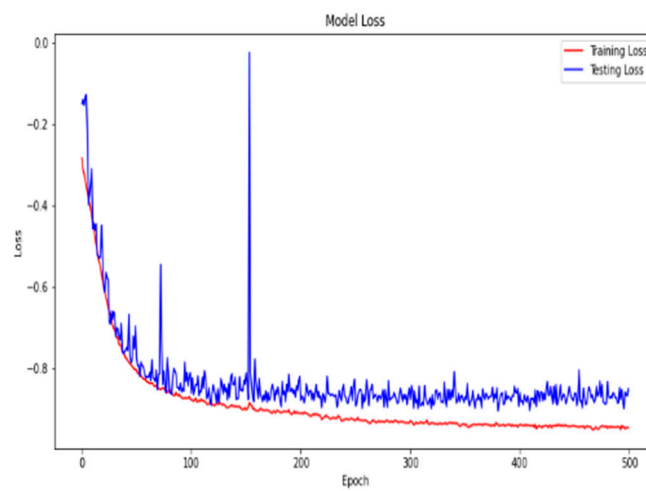
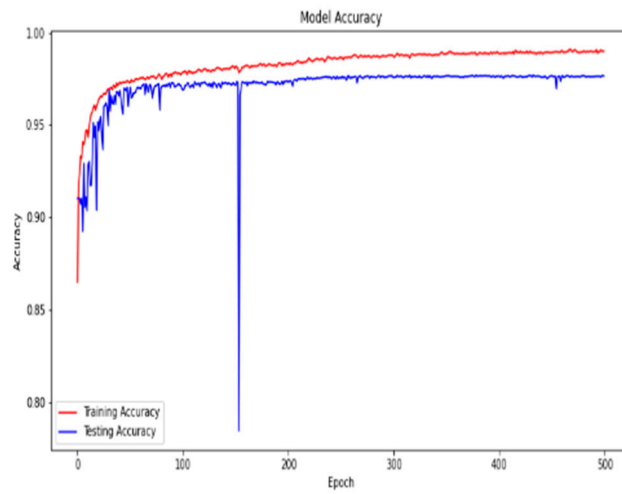


Figure 3. Illustration of implemented U-Net3 architecture.

Figure 4 shows plotting model accuracy and model loss in training and testing from every architecture.



(a) Segnet

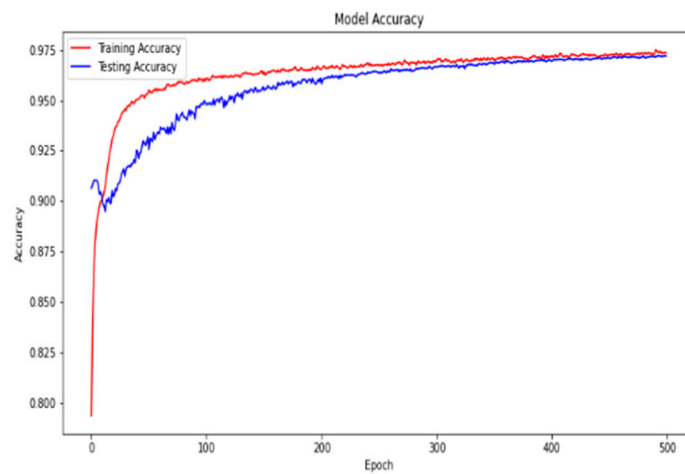
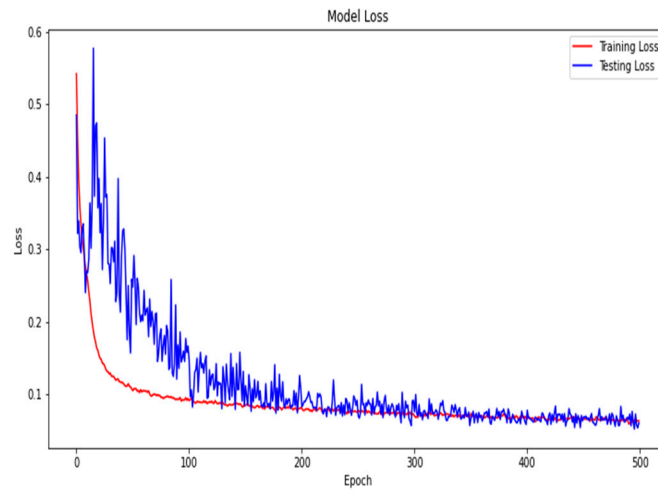
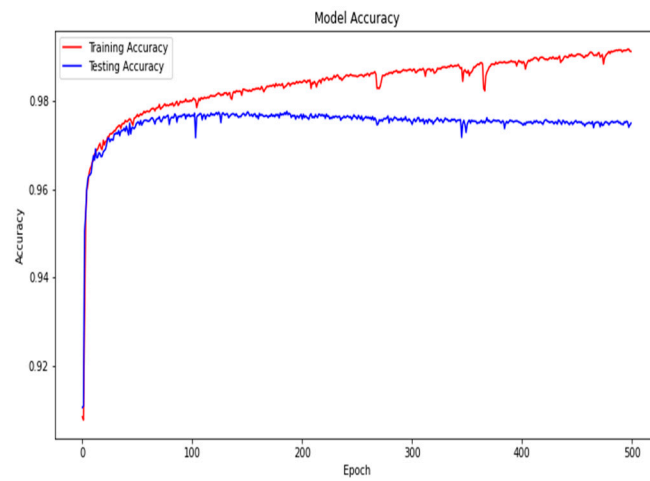


Figure 4. Cont.



(b) ResNet



(c) U-Net

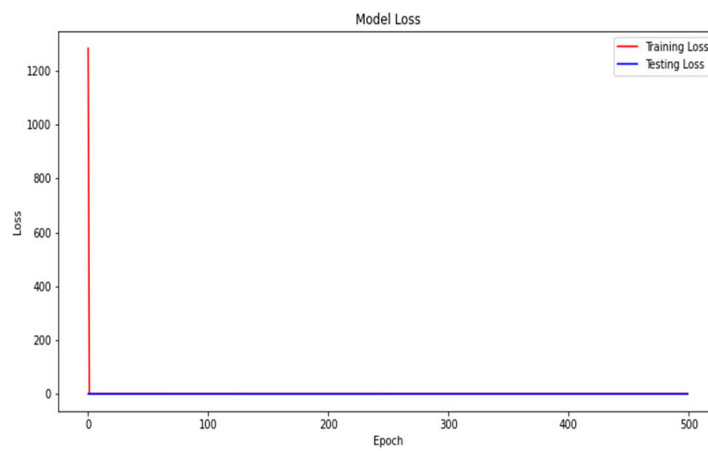
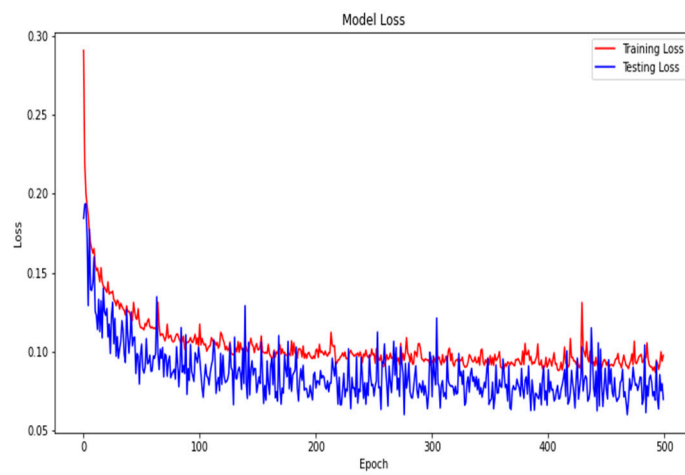
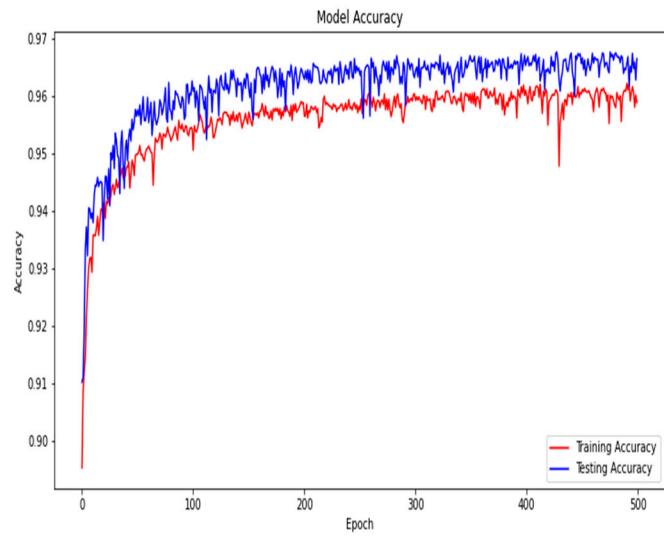


Figure 4. Cont.



(d) U-Net2

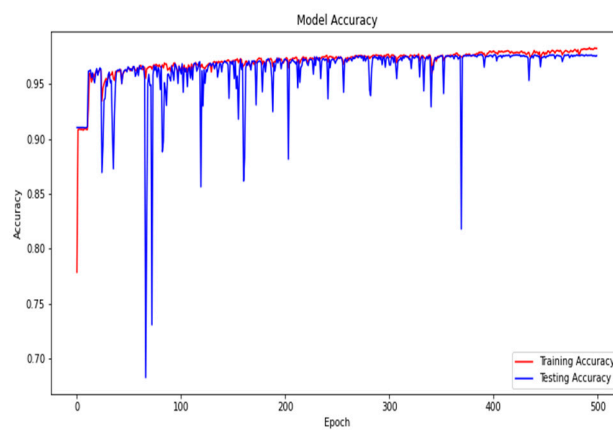
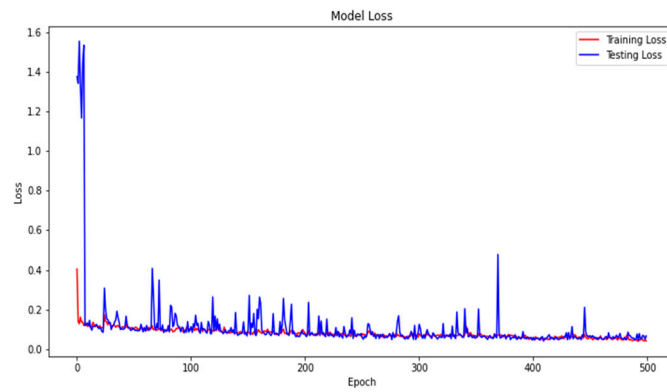
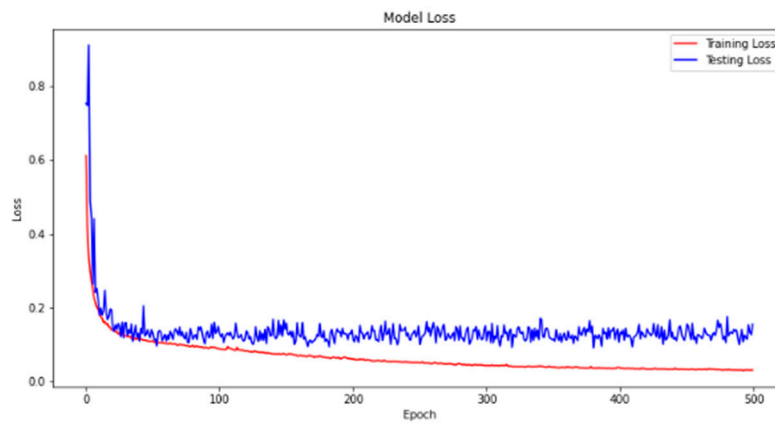
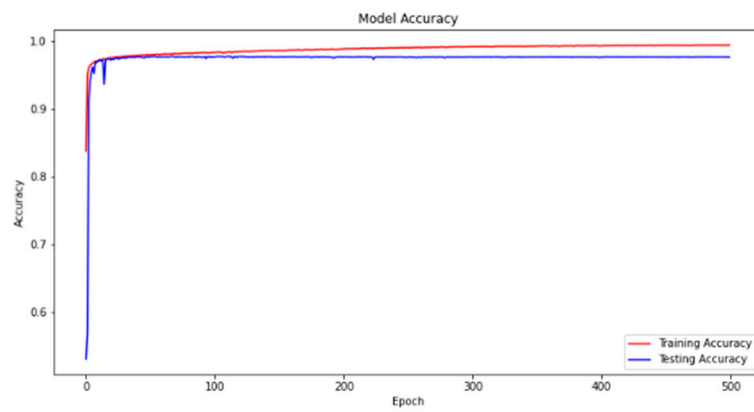


Figure 4. Cont.

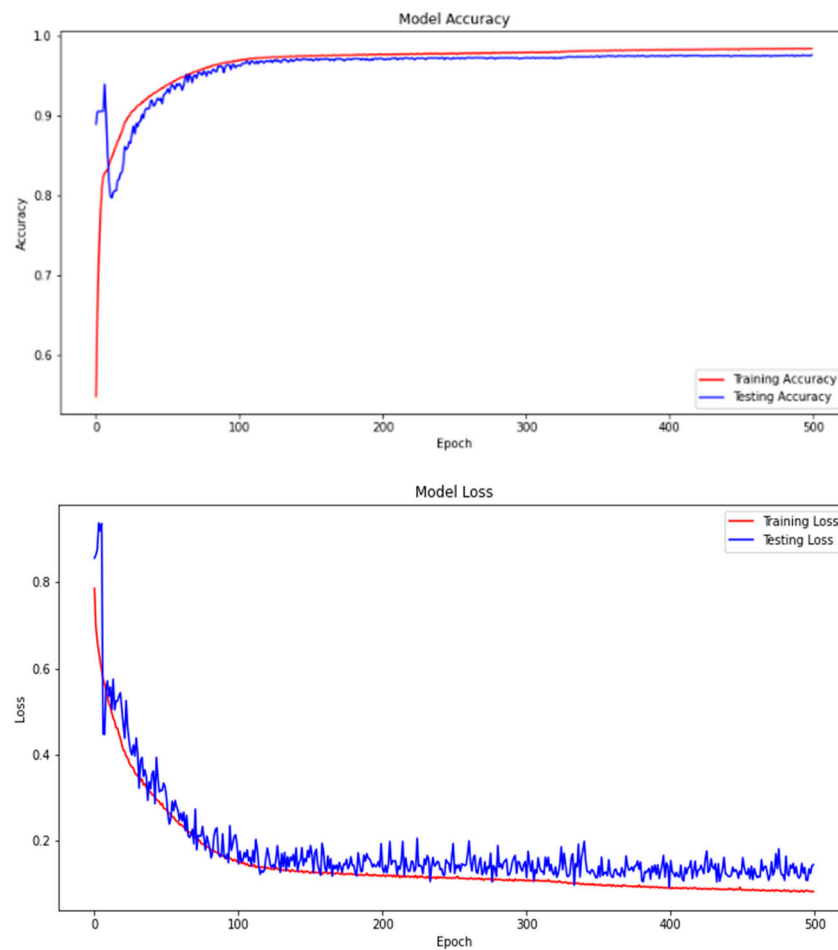


(e) U-Net3



(f) V-Net

Figure 4. Cont.



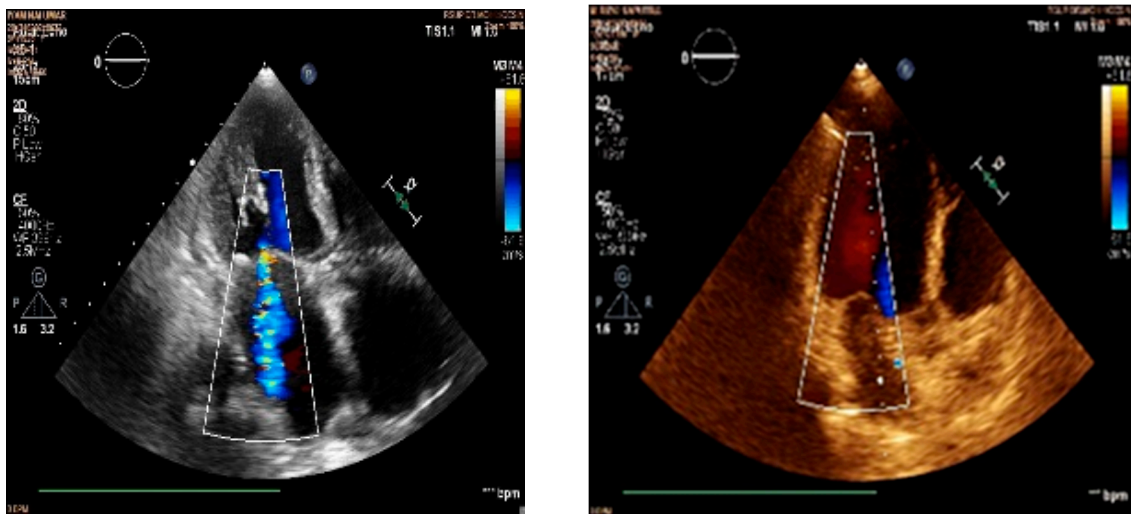
(g) FractalNet

**Figure 4.** Plotting Model loss and Accuracy (a) SegNet; (b) ResNet; (c) U–Net; (d) U–Net2 (e) U–Net3 (f) V-Net, (g) FractalNet, Training(red) vs. testing (blue).

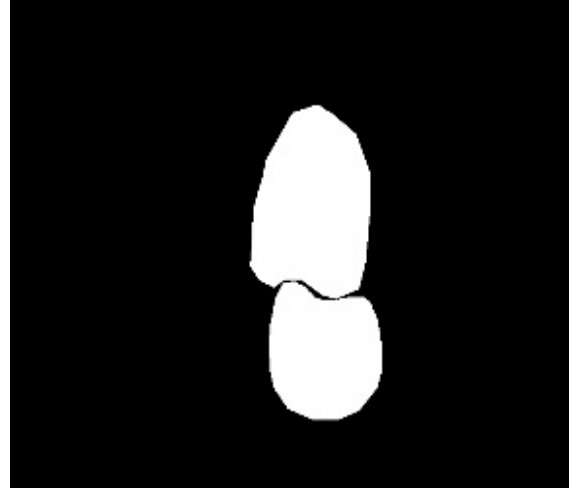
In Table 5, it can be seen that the metric assessment results on the Pixel Accuracy value and the Dice Coefficient on U-Net3 have the highest value compared to other architectures, although other metric evaluations have similar results.

Table 6 shows the results of the entire model by measuring performance using Pixel Accuracy, Intersection over Union, and Dice Coefficient. As can be seen in the table, U-Net3 is the best model, with a Pixel Accuracy value of 97.59%, Intersection over Union value of 86.88%, Mean Accuracy value of 93.46%, Precision value of 85.60%, Recall value of 88.39%, and Dice Coefficient value of 86.58%. We also conducted performance assessments of new data that was not used in training and testing data, unseen data. Although the test results of the U-Net3 model have performance results that are similar to other models; based on the unseen data, the U-Net3 architecture is proven to outperform other architectures, especially on the Dice Coefficient value.

The prediction result of Ground Truth on U-Net3 is close to the result of the Ground Truth, as seen in Figure 5, either in the MR or Normal category.



(a)



(b)



(c)

**Figure 5.** Comparison from (a) Raw Image; (b) Ground Truth; (c) Predicted frame from U-Net3 with 500 epochs.

Table 7 shows that U-Net3 is superior to other architectures because the training time on the U-Net3 model has proven to be faster than other CNN models.

**Table 7.** Performance Training Time.

Architecture	Training Time (s)
Segnet	270.81
ResNet	196.77
U-Net	231.02
U-Net 2	224.53
U-Net 3	194.32
V-Net	357.84
Fractal-Net	295.26

### 3.2. Discussion



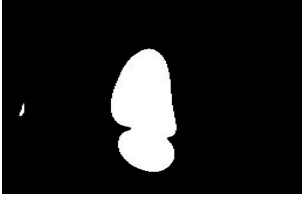

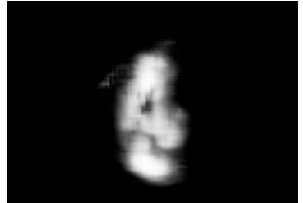

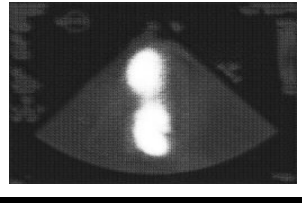

Many studies discuss heart segmentation, but not many discuss heart valve segmentation, in particular, the segmentation of mitral heart valve disease by using the U-Net architecture. As far as we know, this is the first report describing mitral regurgitation segmentation from a color doppler 2D echocardiogram. In another study, an assessment was carried out by assessing the severity of mitral regurgitation to diagnose the severity of mitral regurgitation disease suffered by patients using a deep learning algorithm, namely the convolutional neural network algorithm (Mask R-CNN) in an automated qualitative MR evaluation using color Doppler echocardiography images. The results achieved an average accuracy of around 90 percent. Based on the severity, the classification accuracy was 0.90, 0.89, and 0.91 for mild, moderate, and severe MR [4], respectively, but our proposed method uses the U-Net architecture. Table 8 shows Comparison between ground truth and predicted frame from each architecture.

Based on the predicted ground truth data, it can be seen that the proposed model, U-Net3, succeeded in making predictions that are closest to the ground truth compared to other models because, in other models, the results are not visible. Resnet merges between the top and bottom. U-Net, in the prediction results, shows the upper part is cut off. For U-Net2, the image is also blurred, and is still black in the above image. The V-Net image still looks like the background of the original image, while on fractalNet, in the top and bottom images, it also looks fused.

The research conducted by Nova et al. [14] used U-Net architectures for the automatic segmentation of cardiac septal defects. They proposed a CNN-based U-Net to automatically analyze cardiac chamber segments to detect abnormalities (holes) in the cardiac septum. In this study, segmentation was performed on atrial septal defects (ASD), ventricular septal defects (VSD), septal defects (AVSD), and a normal heart. The results showed that the proposed method can produce superior performance in detecting cardiac septal defects. The segmentation model for the four classes resulted in a Pixel Accuracy of 99.15%, an average Intersection over Union (IoU) of 94.69%, and F1 scores of 94.88%, respectively. The research also compared the proposed models of U-Net and V-Net architecture. The result of the accuracy of the prediction contour comparison for U-Net was 99.01%, while V-Net was 93.70%. It can be concluded that the U-Net model architecture has higher accuracy than V-Net. Meanwhile, in the research that we propose, we also use the U-Net architecture that has been developed, namely U-Net3. As for the results, although we have pixel accuracy of 97.62%, IoU of 86.93%, and F1-Score of 86.51, which is slightly smaller, the difference lies in the object segmentation, the architectural model, and the epoch used. We propose segmentation of mitral regurgitation with the U-Net3 architectural model and also compare it with six other architectures, namely Vnet, Segnet, Resnet, FractalNet, U-Net, and U-Net2. We currently use 500 epochs, while this study proposes segmentation of abnormalities (holes) in the cardiac septum, the architectural model used is U-Net, which is compared to V-Net architecture and 1000 epochs of epoch.



**Table 8.** Comparison between ground truth and predicted frame from each architecture.

Ground Truth	Arsitektur	Predict Ground Truth
	Segnet	
	ResNet	
	U-Net	
	U-Net 2	
	U-Net 3	
	V-Net	
	Fractal-Net	

In a study conducted by Rachmatullah et al. [21], they proposed the U-Net architecture for automatic segmentation on the fetal heart image. Regarding the data in the segmentation of ultrasound images, as many as 519 images of the fetal heart were obtained from three videos. In that paper, the combination of U-Net and Otsu threshold resulted in a fairly good performance; 99.48% Pixel Accuracy, 94.92% IoU, and 0.21% error rate. This study discussed fetal heart, which was almost the same as previous studies, whereas we discussed mitral disease regurgitation, which is still rarely discussed. There were 519 images used in that research, while we used 923 images.

In a previous study by Diniz et al., the U-Net architecture model with Concatenation Block (Concat U-Net) managed to get good results for cardiac segmentation. The results of the study reached a Dice Coefficient of 87.95%. This study discussed CT Scan Heart, whereas we discussed mitral disease regurgitation and produced a Dice Coefficient of 86.58%. Although the results were not too different, our research is still rarely carried out and we have also made ground truth predictions, as shown in Table 9.

In this study, the proposed U-Net3 architecture can segment regurgitation mitral valves and normal heart valves in 4-chamber heart data, which has a Pixel Accuracy value of 97.59%, Intersection over Union value of 86.88%, Mean Accuracy value of 93.46%, Precision value of 85.60%, Recall value of 88.39%, and Dice Coefficient value of 86.58%. With a high accuracy value, the prediction of the accuracy is close to the original image. Although there has been no segmentation research on heart valves, the results show that the U-Net3 architecture is quite superior because it has a high accuracy value. The Dice Coefficient value is 86.58%. In addition to the Dice Coefficient value on unseen data, the U-Net3 model also has a training time that has proven to be faster than other CNN models. In this study, U-Net3 has been a proven prediction in segmentation using the proposed datasets. U-Net3 has the highest metric evaluation in Pixel Accuracy, Intersection Over Union, Recall, And Dice Coefficient. In addition to the Dice Coefficient value in the unseen data in Table 6, the U-Net3 model also has training time which is proven to be faster than other CNN models, as seen in Table 7.

There are many approaches used to segment various medical objects using CNN. Table 5 shows a comparison of the results of several previous studies and the current approach.

**Table 9.** Comparison of other methods with different data for segmentation.

No	Author	Year	Study	Architecture	Dice Coefficient
1	Nova et al. [14]	2021	Fetal Heart Echocardiography	U-Net	94.88%
2	Rachmatullah et al. [21]	2021	Fetal Heart Echocardiography	U-Net and Otsu threshold	87.95%
3	Diniz et al. [22]	2021	CT Scan heart	U-Net with Concat U-Net	96.71%
4	Proposed	2022	Mitral Valve Echocardiography	U-Net3	86.58%

#### 4. Conclusions

In this paper, we proposed a segmentation model using a CNN-based U-Net3 architecture that successfully explains the segmentation of heart valve disease mitral regurgitation.

We measured and evaluated the performance of the proposed model using the parameters of Pixel accuracy, IoU, Mean Accuracy, Precision, Recall, and F1 Scores, obtaining score values of 97.59%, 86.98%, 93.46%, 85.60%, 88.39%, and 86.58%, respectively. The best performance is obtained on the U-Net3 architecture with a batch size of 64 and the loss function of binary cross entropy. We also compared the proposed model with six other architectures, namely SegNet, ResNet, FractalNet, U-Net and U-Net2, and tested all seven architectures using unseen data, which is new data that was not used during training. From the results of the experiment, it can be concluded that U-Net3 is the best predictor in predicting ground truth that is close to the original image. In future work, we will increase

the number of patients even more, and also add an epoch of 1000 epochs to produce even better ground truth.

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

**Funding:** This research received BPPDN Scholarship, Indonesia Government.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data is contained within the article.

**Acknowledgments:** This research was funded by Universitas Bina Darma.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Callow, A.D. Cardiovascular disease 2005—The global picture. *Vasc. Pharmacol.* **2006**, *45*, 302–307. [[CrossRef](#)] [[PubMed](#)]
2. Mozaffarian, D.; Benjamin, E.J.; Go, A.S.; Arnett, D.K.; Blaha, M.J.; Cushman, M.; Das, S.R.; de Ferranti, S.; Després, J.-P.; Fullerton, H.J.; et al. Heart Disease and Stroke Statistics-2016 Update: A Report From the American Heart Association. *Circulation* **2016**, *133*, e38–e360. [[CrossRef](#)] [[PubMed](#)]
3. Gumireddy, S.R.; Katayama, M.; Chaliki, H.P. A Case of Severe Mitral Valve Regurgitation in a Patient with Leadless Pacemaker. *Case Rep. Cardiol.* **2020**, *2020*, 5389279. [[CrossRef](#)] [[PubMed](#)]
4. Zhang, Q.; Liu, Y.; Mi, J.; Wang, X.; Liu, X.; Zhao, F.; Xie, C.; Cui, P.; Zhang, Q.; Zhu, X. Automatic Assessment of Mitral Regurgitation Severity Using the Mask R-CNN Algorithm with Color Doppler Echocardiography Images. *Comput. Math. Methods Med.* **2021**, *2021*, 2602688. [[CrossRef](#)] [[PubMed](#)]
5. Mayasari, N.M.; Anggrahini, D.W.; Mumpuni, H.; Krisdinarti, L. Incidence of Mitral Valve Prolapse and Mitral Valve Regurgitation in Patient with Secundum Atrial Septal Defect. *Acta Cardiol. Indones.* **2015**, *1*, 5–7.
6. Esteva, A.; Robicquet, A.; Ramsundar, B.; Kuleshov, V.; Depristo, M.; Chou, K.; Cui, C.; Corrado, G.; Thrun, S.; Dean, J. A guide to deep learning in healthcare. *Nat. Med.* **2019**, *25*, 24–29. [[CrossRef](#)] [[PubMed](#)]
7. Popescu, D.; El-Khatib, M.; El-Khatib, H.; Ichim, L. New Trends in Melanoma Detection Using Neural Networks: A Systematic Review. *Sensors* **2022**, *22*, 496. [[CrossRef](#)] [[PubMed](#)]
8. Minaee, S.; Boykov, Y.Y.; Porikli, F.; Plaza, A.J.; Kehtarnavaz, N.; Terzopoulos, D. Image Segmentation Using Deep Learning: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 3523–3542. [[CrossRef](#)] [[PubMed](#)]
9. Skourt, B.A.; El Hassani, A.; Majda, A. Lung CT image segmentation using deep neural networks. *Procedia Comput. Sci.* **2018**, *127*, 109–113. [[CrossRef](#)]
10. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics); Springer: Cham, Switzerland, 2015; Volume 9351, pp. 234–241. [[CrossRef](#)]
11. Siddique, N.; Paheding, S.; Elkin, C.P.; Devabhaktuni, V. U-net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access* **2021**, *9*, 82031–82057. [[CrossRef](#)]
12. Zhang, Q.; Cui, Z.; Niu, X.; Geng, S.; Qiao, Y. Image Segmentation with Pyramid Dilated Convolution Based on ResNet and U-Net. In *International Conference on Neural Information Processing*; Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics); Springer: Cham, Switzerland, 2017; Volume 10635, pp. 364–372. [[CrossRef](#)]
13. Liciotti, D.; Paolanti, M.; Pietrini, R.; Frontoni, E.; Zingaretti, P. Convolutional Networks for Semantic Heads Segmentation using Top-View Depth Data in Crowded Environment. In *Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR)*, Beijing, China, 20–24 August 2018; pp. 1384–1389. [[CrossRef](#)]
14. Nova, R.; Nurmaini, S.; Partan, R.U.; Putra, S.T. Automated image segmentation for cardiac septal defects based on contour region with convolutional neural networks: A preliminary study. *Inform. Med. Unlocked* **2021**, *24*, 100601. [[CrossRef](#)]
15. Kalane, P.; Patil, S.; Patil, B.; Sharma, D.P. Automatic detection of COVID-19 disease using U-Net architecture based fully convolutional network. *Biomed. Signal Process. Control* **2021**, *67*, 102518. [[CrossRef](#)] [[PubMed](#)]
16. Milletari, F.; Navab, N.; Ahmadi, S.-A. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In *Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV)*, Stanford, CA, USA, 25–28 October 2016; pp. 565–571. [[CrossRef](#)]
17. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [[CrossRef](#)] [[PubMed](#)]
18. Larsson, G.; Maire, M.; Shakhnarovich, G. FractalNet: Ultra-Deep Neural Networks without Residuals. *arXiv* **2016**, arXiv:1605.07648.

19. Zhuang, J. LadderNet: Multi-path networks based on U-Net for medical image segmentation. *arXiv* **2018**, arXiv:1810.07810.
20. Benjdira, B.; Ammar, A.; Koubaa, A.; Ouni, K. Data-efficient domain adaptation for semantic segmentation of aerial imagery using generative adversarial networks. *Appl. Sci.* **2020**, *10*, 1092. [[CrossRef](#)]
21. Rachmatullah, M.N.; Nurmaini, S.; Sapitri, A.I.; Darmawahyuni, A.; Tutuko, B.; Firdaus, F. Convolutional neural network for semantic segmentation of fetal echocardiography based on four-chamber view. *Bull. Electr. Eng. Informatics* **2021**, *10*, 1987–1996. [[CrossRef](#)]
22. Diniz, J.O.B.; Ferreira, J.L.; Cortes, O.A.C.; Silva, A.C.; de Paiva, A.C. An automatic approach for heart segmentation in CT scans through image processing techniques and Concat-U-Net. *Expert Syst. Appl.* **2022**, *196*, 116632. [[CrossRef](#)]