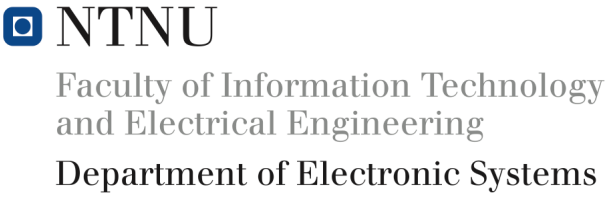
Auto-strain master thesis

Written by
Yohann Jacob Sandvik

Master thesis EMNEKODE

Supervised by
Lasse Løvstakken



Department of electronic systems
Faculty of Information Technology and Electrical Engineering
Norwegian University of Science and Technology
June 9, 2020

Abstract

This is the abstract.

Acknowledgements

These are my acknowledgements.

Contents

List of Abbreviations

BMI Body Mass Index. 4, 13

EF Ejection Fracture. 4, 14, 16, 17, 20

GLS Global Longitudinal Strain. 4, 14, 16–20

ML Machine Learning. 19

RLS Regional Longitudinal Strain. 19, 20

RNN Recurrent Neural Network. 20

TSC Time-series clustering. 20

List of Figures

List of Tables

Chapter 1

Introduction

This is the introduction.

1.1 Motivation

This will be the section on the motivation for the assignment.

1.2 Objective

This will be the section where i outline the objective of the assignment.

1.3 Structure of Thesis

Here the outline for the rest of the assignment will be given.

Chapter 2

Myocardial Imaging and Echocardiography

This will be a kind of theory section about echocardiography, and strain imaging.

2.1 Basic Cardiology

2.2 Introduction to Echocardiography

2.3 Myocardial Strain

Chapter 3

Machine Learning Theory

This section will act as a theory section for the machine learning models used.

Chapter 4

Review of The Literature

This chapter will contain the review of the literature.

Chapter 5

Data Exploration

In this chapter the variability, distribution and type of data used in the assignment will be explored. The exploration is divided into three sections corresponding to the three main groups of variables: The *patient meta-data*, the *input variables* and the *target variables*. The *meta-data* is the data about the patients which is not used in the classification models, but can be used to give a description of the patient demographich which makes up the dataset. The *input variables* are the variables that are inputed into the machine learning models in order to train them, and later used to make predictions about the patients' *target variables*. The target variables are then variables that the models will be trained to predict. Target variables are used both in training to correct erroneus predictions that models make during training, and to evaluate the accuracy of the model after training.

5.1 Patient Meta-data

The patient meta-data that will be considered in this section are age, gender, Body Mass Index (BMI) and blood pressure.
Figure 5.1 shows the patient distributions with regard to age, gender and BMI. As evident from the figure the patients that make up the dataset is made up of 138 males and 57 females. The majority of the patients are in the age group 60-80 years with a number of patients in the range 80-90 years (AGE SECTION SUBJECT TO CHANGE). The BMI distribution of patients is centered around 26 kg/m^2. Figure 5.2 shows the joint distribution of systolic and diastolic blood pressure among the patients.

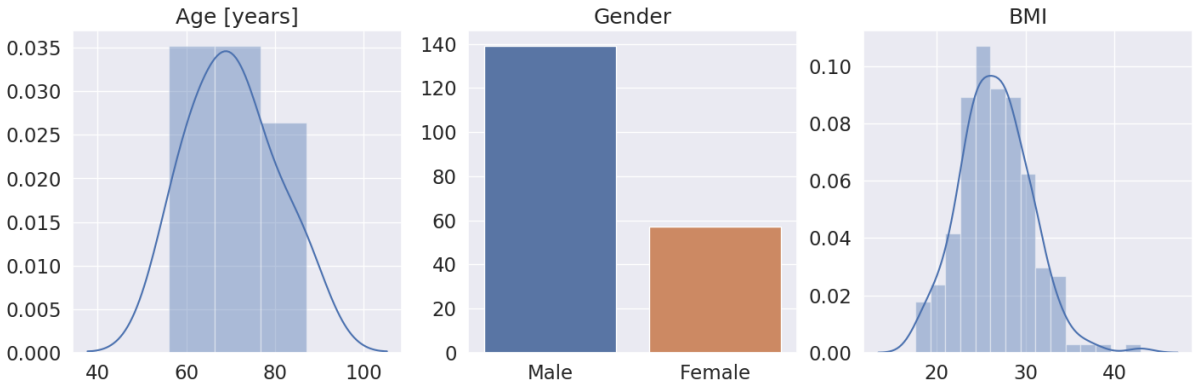


Figure 5.1: Distribution of age, gender and BMI.

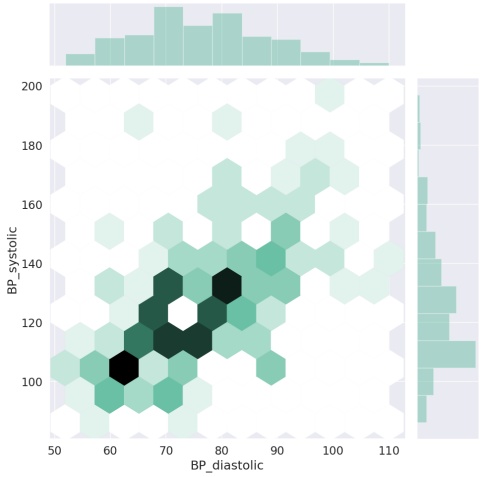


Figure 5.2: A joint distribtion plot of systolic and diastolic blood pressure of the patients.

5.2 Input variables

As mentioned earlier in section REF the different machine learning models that will be applied will apply two types of input data, time-series data in the form of longitudinal strain curves, and point-values in the form of peak systolic global longitudinal strain and patient EF.

5.2.1 Peak values

As mentioned in section REFERENCE EF values below 40-50% is regarded as unhealthy with regard to probability of heart failure. Keeping that in mind, one should note that the distribution of EF values among the patients shown in figure 5.3 is centered at approximately 40% with tails going as low as 8% and as high as 70%. Figure 5.4 shows the distribution of peak systolic GLS values, four the three different views. As evident from the figure, the values are centered around -12.5 with tails going as low as -29, and as high as -2.5.

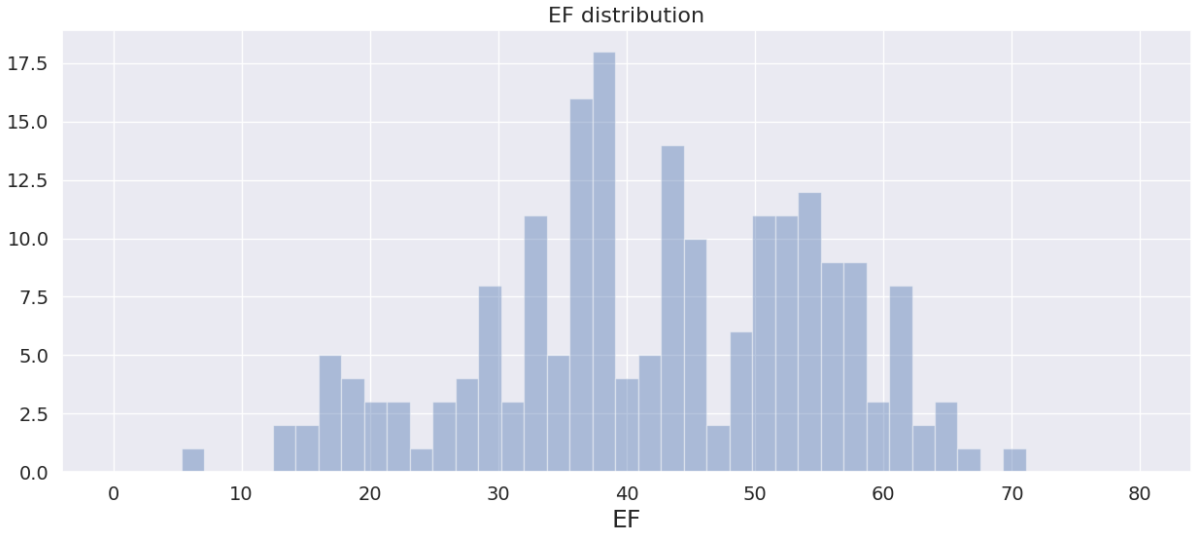


Figure 5.3: Distribution of patient EF values.

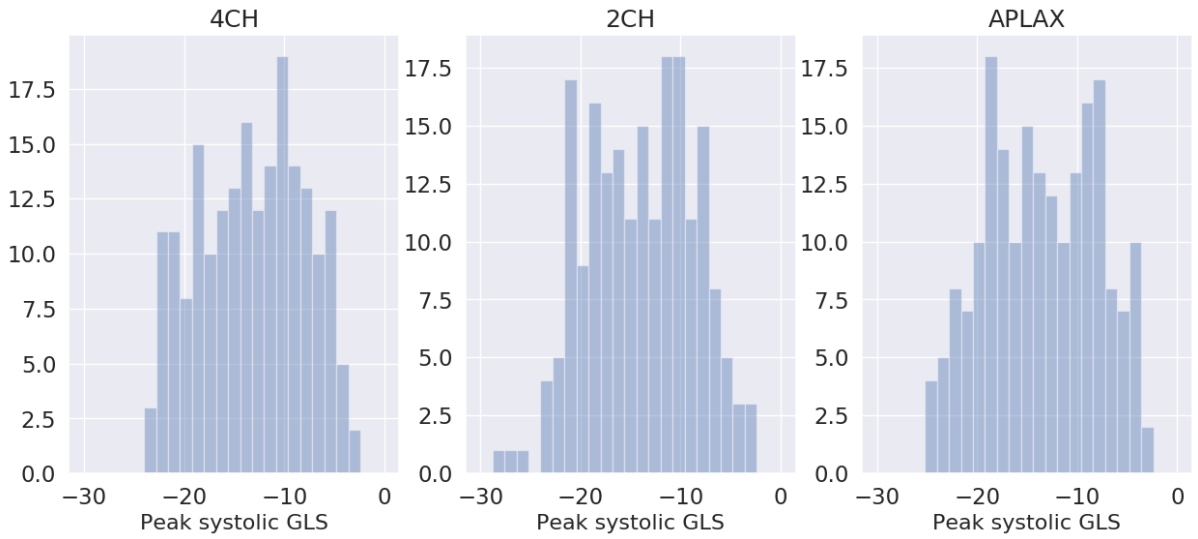


Figure 5.4: Distribution of peak systolic global longitudinal strain.

5.2.2 Strain curves

Figure 5.5 shows what a typical set of strain curves look like for a patient. Only the six regional strain curves, and the one global strain curve from the 4CH view have been included as they are fairly similar across the different views. Since the data from the different patients have been taken at different times, and possibly with different ultrasound machines factors such as number of samples per strain curve, and the frame rate of the particular ultrasound machine during an examination. Each strain curve has a standardized length of one heart cycle, due to this different curves have different number of samples. Figure 5.6 shows the distribution of frame rates, and number of samples among the total number of strain curves.

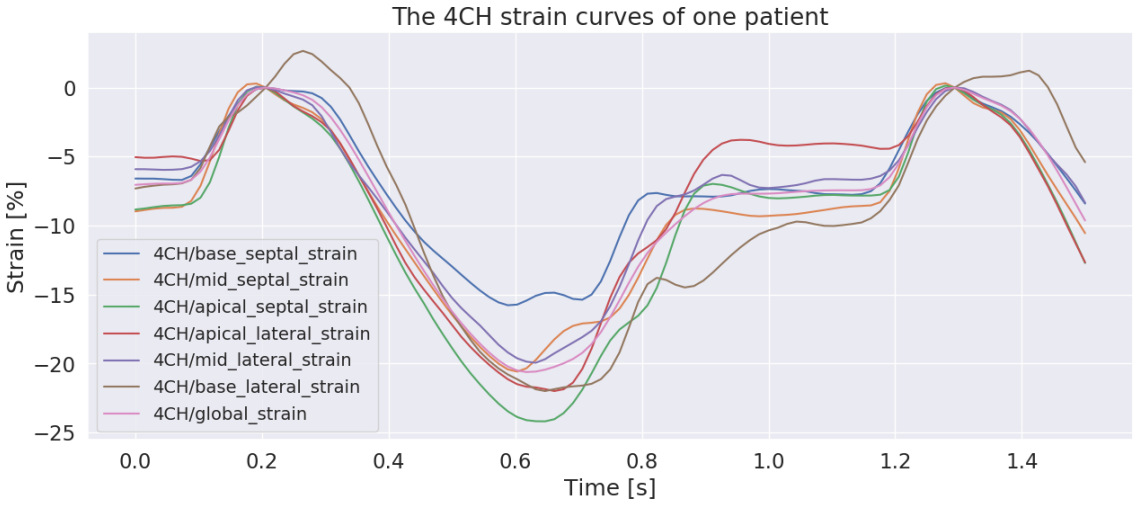


Figure 5.5: Plot of the global and regional longitudinal strain curves of one patient in the 4CH view.

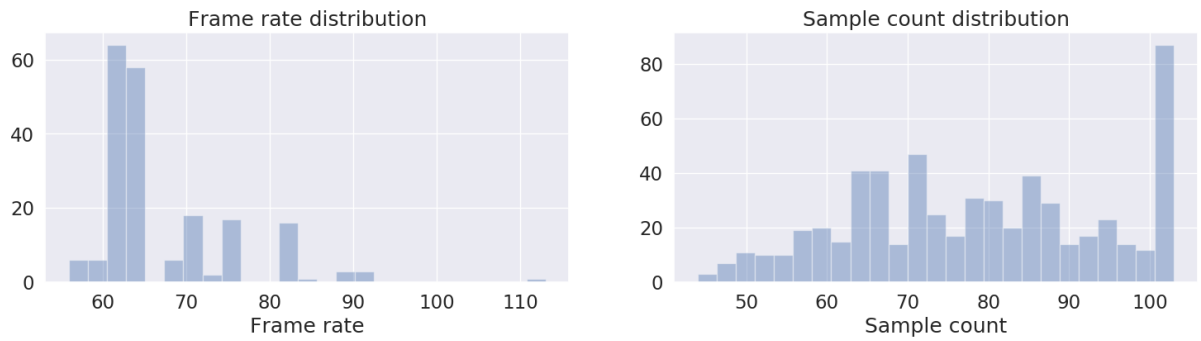


Figure 5.6: Distribution of the frame rate used in the ultrasound imaging used to obtain the strain curves (left), and sample count of the different strain curves (right).

5.3 Target variables

Figure 5.7 shows the distribution of heart failure among patients (left), and the distribution of different indications (right). Since the dataset has approximately as many patients with a heart failure diagnosis as without, it can be considered balanced in that regard. With regard to the different patient diagnosises, their rate of occurance can is not uniform in this dataset. The control group of healthy individuals consists of 31 patients. The groups of patients with STEMI, and NSTEMI indications consist of 60 and 39 patients respectively. Finally, the group of patients with heart failure, but with a non-stemic indication (labelled OTHER in left barplot in figure 5.7) consists of 69 patients.

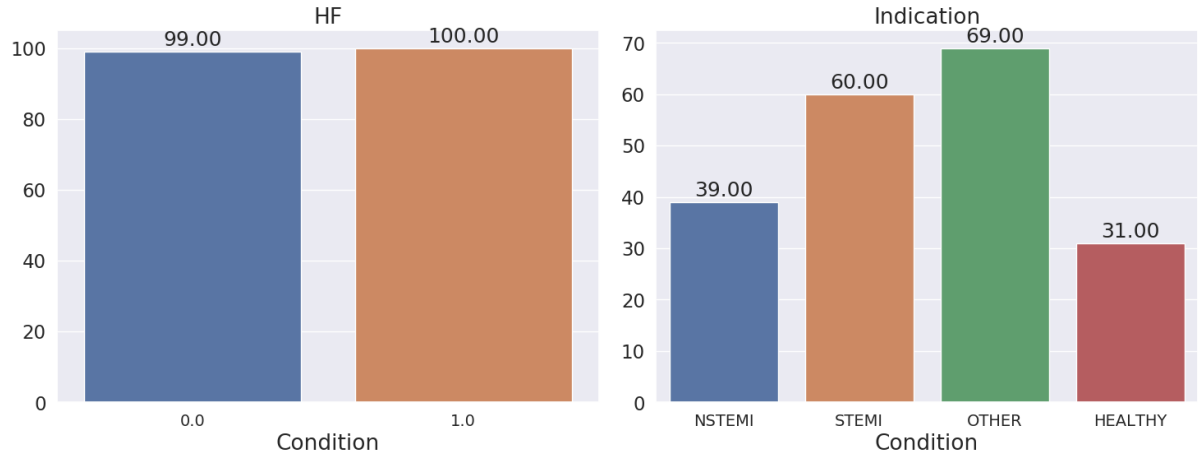


Figure 5.7: The distribution of heart failure and different indications within patients.

To illustrate the diagnostic power of EF, and peak strain values figure 5.8 shows the distribution of EF for patients with and without heart failure (left), and the distribution of EF for patients in the control group and the other patients (right). Figure 5.9 shows the distribution of peak systolic GLS values for patients with and without heart failure, and figure 5.10 shows the distribution of peak systolic GLS values for patients in the control group and the rest of the patients. From the samples used to produce the left plot in figure 5.8 and figure 5.9 it seems as though the heart failure patients are more separable with the EF values than with the GLS values. With regard to separability of patients with diagnoses and patients in the control group it seems as though the right plot in figure 5.8, and figure 5.10 follows the same distribution as

the heart failure patients. However, it is hard to make an evaluation on this since the sample size of the control group is much smaller than the group of diagnosed patients.

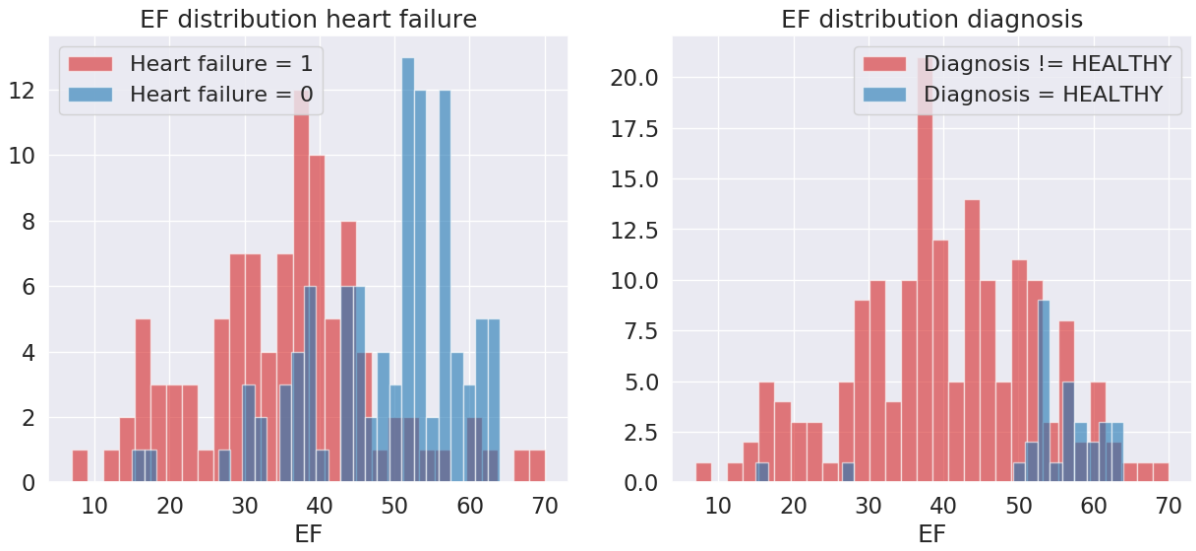


Figure 5.8: Distribution of EF for patients with and without heart failure (left), and distribution of EF for patients in the control group, and patients with a diagnosis.

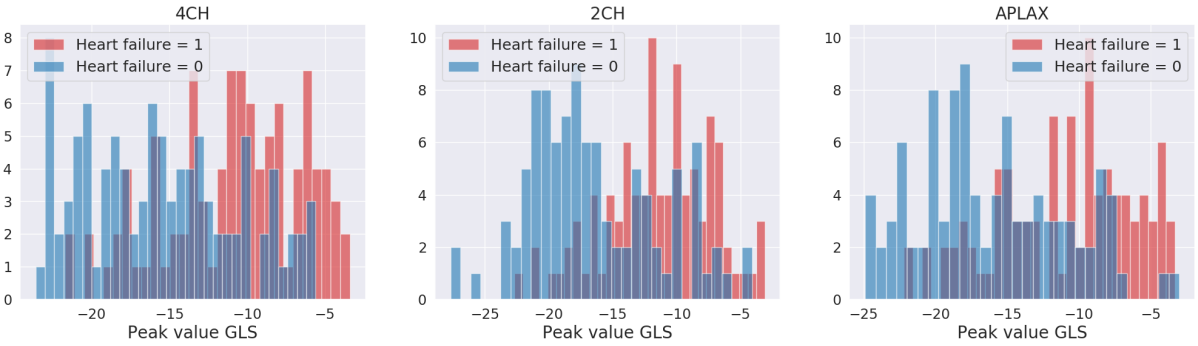


Figure 5.9: Distribution of GLS for patients with and without heart failure.

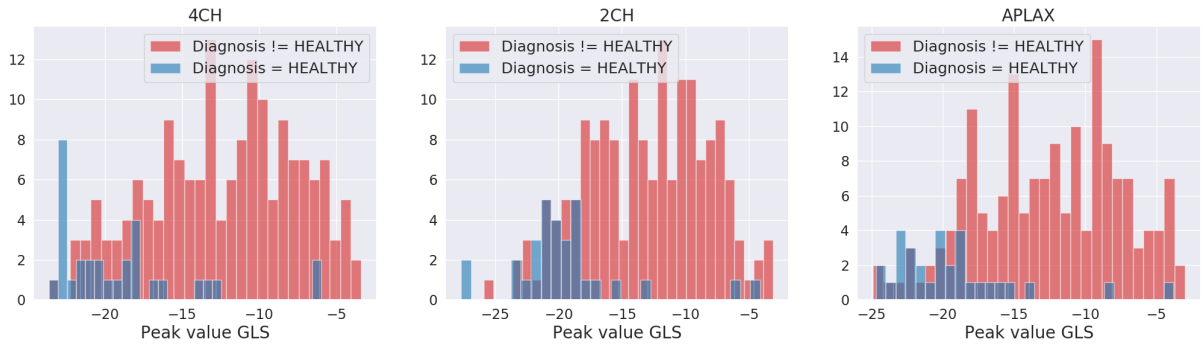


Figure 5.10: Distribution of GLS for patients in the healthy control group, and the other patients.

Figure 5.11 shows the distribution of the different segment indications, for all the left ventricle segments of all the patients in the dataset. Since the occurance of indications other that "normal" and "hypokinetic" are very rare, the occurance axis has been used as logarithmic. The imbalance of segment-indication labels illustrated in figure 5.11 means that it will challenging for any statistical model to perform well in the classes with low occurance. To counteract this one can change the taxonomy of the labels such that the classification problem becomes binary with the labels *Normal* and *Not normal*. The dataset is then fairly evenly distributed with 1695 *Normal* labels and 1818 *Not normal* labels.

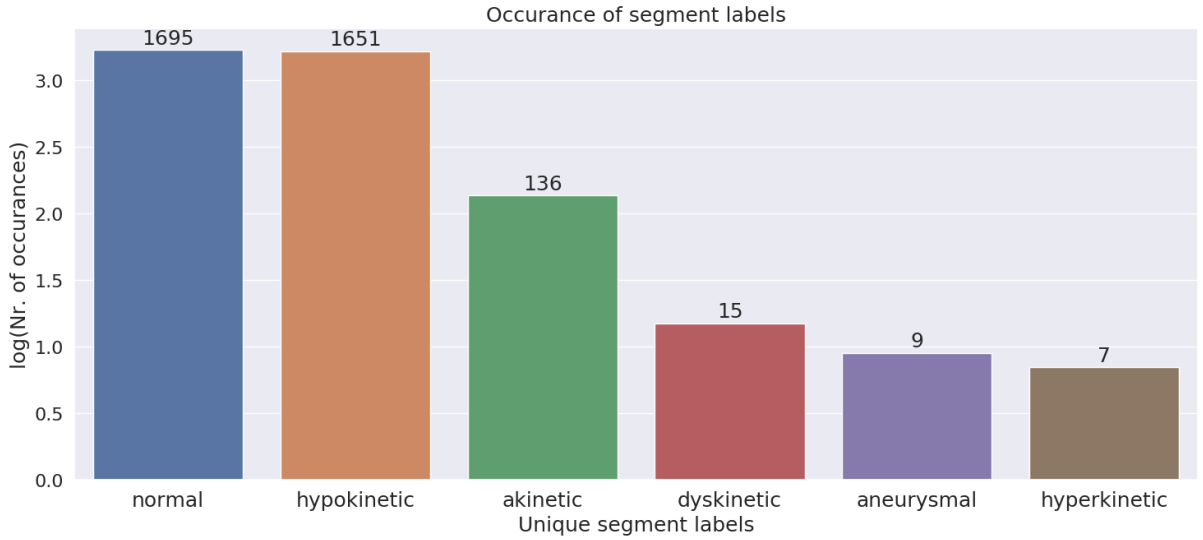


Figure 5.11: Distribution segment indication labels.

Chapter 6

Method

6.1 Models

This is the section where we detail the specific models used.

6.1.1 Time-series clustering

6.1.2 Peak-value clustering

6.1.3 Recurrent Neural Network

6.1.4 Supervised Peak-value Classifiers

6.2 Description of The Datasets

Since the different ML models detailed in chapter REFERENCE require different types of input data the, datasets have been divided into two main categories: The peak-value datasets and the time-series datasets.

6.2.1 Time-series Datasets

Nr	Input variables	Shape
1	Single RLS curves	$(3600, 1)$
2	RLS curves	$(200, 18)$
3	GLS curves	$(200, 3)$
4	Strain curves	$(200, 21)$

Table 6.1: Time-series datasets. The "Shape" parameter is indicates: (Number of objects in the dataset, Number of curves in each individual object). The curve length is not included in the shape parameter because it differs for different curves.

Table 6.1 shows the different time-series datasets that will be used. All the datasets except *Single RLS curves* will be used to predict whether or not the patient is diagnosed, and whether the patient has heart failure. Recall that the different diagnosises are described in section REFERENCE, and there occurance rate are illustrated in figure 5.7. *Single RLS curves* will be used to

predict the segment indications shown in figure **??** and described in section REFERENCE. The point of classifying individual segments of a patients left ventricle is that if a single segment is found to be *not normal*, this would also mean that the patient can be considered as *not healthy*. As mentioned in the description of table 6.1 the "Shape" parameter shows how many objects each dataset has, and how many curves are associated to each object. Since each ultrasound examination takes ultrasound inspections from three views (four chamber, two chamber, and APLAX chamber), each patient has three views to estimate a GLS curve from. Since each GLS curve, also can be divided into six RLS curves, there is a total of 21 strain curves per patient. Since each patient has 18 RLS curves, there are $18 \times 200 = 3600$ curves that make up dataset number 1. Both the RNN, and the TSC model are applied on the datasets listed in table 6.1,

6.2.2 Peak-value Datasets

Nr	Input variables	Shape
1	Single peak systolic RLS values	$(3600, 1)$
2	Peak systolic RLS values	$(200, 18)$
3	Peak systolic GLS values	$(200, 3)$
4	Peak systolic strain values	$(200, 21)$
5	Peak systolic RLS, and EF values	$(200, 19)$
6	Peak systolic GLS, and EF values	$(200, 4)$
7	Peak systolic strain, and EF values	$(200, 22)$

Table 6.2: Peak-value datasets. The "Shape" parameter is indicates: (Number of objects in the dataset, Number of dimensions of each individual object).

Table 6.2 shows the different peak-value datasets. All the datasets with exception of *Single peak systolic RLS values* will be used to predict the diagnosis of patients, and whether the patient has heart failure. *Single peak systolic RLS values* is also the only peak-value dataset that is not suited for clustering, since a minimum of two dimensions is required to cluster a point-value dataset. The reason that there are more peak-value datasets than there are time-series datasets, is that the peak-value version of three datasets in table 6.1 have been combined with EF to determine whether a combination of peak systolic strain, and EF can have a higher predictive power than strain alone.

Chapter 7

Results

In this chapter the results will be presented in the form of three case studies. Each case study will focus on a single target variable, and aims to find which model group performs best at predicting the target variable in question. Recall that the three target variables that will be considered in this thesis are: Heart failure, patient diagnosis, and the indication of individual left ventricle segments. As mentioned earlier in the chapter, four model groups will be tested. The case studies will first deal with each model group individually, where variants of the models with different hypermarameters will be tested on the different datasets. Then, the best performing model within each model group will be used to compare the four model groups. The supervised models will be assessed with the metrics: accuracy, sensitivity, specificity and DOR. The clustering methods evaluated at two cluster centers will be assessed with the same methods as the supervised models. The clustering methods evaluated at two to nine cluster centers will also be assessed with ARI to determine whether the models evaluated at a higher number of cluster centers could fit the data better.

7.1 Case Study: Heart Failure

7.1.1 Time-series Clustering

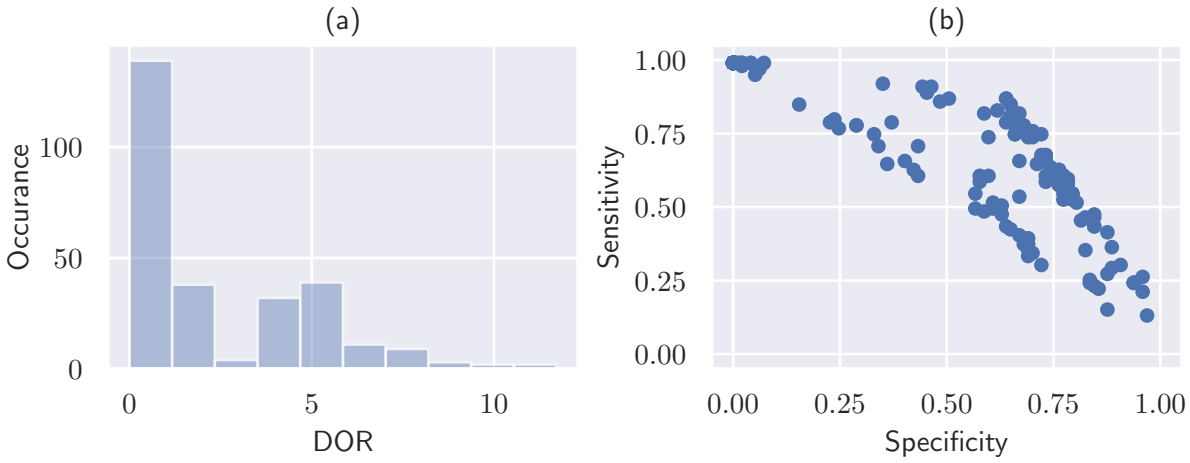


Figure 7.1: (a) Distribution plot of DOR of all TSC methods evaluated at two cluster centers when applied to classify heart failure. (b) Scatter plot of the same methods sensitivity, and specificity.

Dataset-Method	Accuracy	Sensitivity	Specificity	DOR
gls/2CH/regular/centroid/2	0.76	0.87	0.64	11.72
gls/2CH/scaled/centroid/2	0.76	0.87	0.64	11.72
gls/2CH/regular/average/2	0.75	0.85	0.65	10.38
gls/2CH/scaled/average/2	0.75	0.85	0.65	10.38
gls-rls/2CH/scaled/ward/2	0.74	0.82	0.67	9.14

Table 7.1: The accuracy, DOR, sensitivity and specicity scores of the five best performing two-cluster-center TSC methods in terms of DOR, at detecting heart failure. The **Dataset-Method** column indicates *Dataset used/ View used/ Type of preprocessing used/ Linkage criteria of method/ Number of cluster centers.*

Figure 7.1a shows that the DOR is close to zero for many of the two-cluster-center methods, However, the best performing methods are able to acheive a DOR above ten, these methods are listed in table 7.1. From the scatterplot in figure 7.1b one can see that the distribution of sensitivity, and specificity are quite widespread. Sensitivity and specificity scores range from 0 to 1. Common to the top 18 methods in terms of DOR is that they all use data from a single view, and 2CH is the only view that is represented among the five methods with highest DOR. What else is worth noting is that almost all the methods using normalization or z-normalization as preprocessing score below the methods that use scaling, or no preprocessing at all. These observations can be confirmed from the table10.1 in the appendix. From table 7.1 one can see that the two best-performing methods in terms of DOR received the exact same score in all metrics. *gls/2CH/regular/centroid/2*, and *gls/2CH/scaled/centroid/2* differ only in the way of preprocessing, the former does not preprocess the curves before clustering, and the latter uses scaling. However, for these two cases preprocessing did not matter as they have the exact same cluster assignments as well.

Dataset-Method	ARI
gls/2CH/regular/centroid/2	0.25
gls/2CH/scaled/centroid/2	0.25
gls/2CH/scaled/centroid/3	0.24
gls/2CH/regular/centroid/3	0.24
gls/2CH/scaled/average/2	0.24

Table 7.2: The five highest ARI scores attained when applying TSC for detecting heart failure. The **Dataset-Method** column indicates *Dataset used/ View used/ Linkage criteria of method/ Number of cluster centers.*

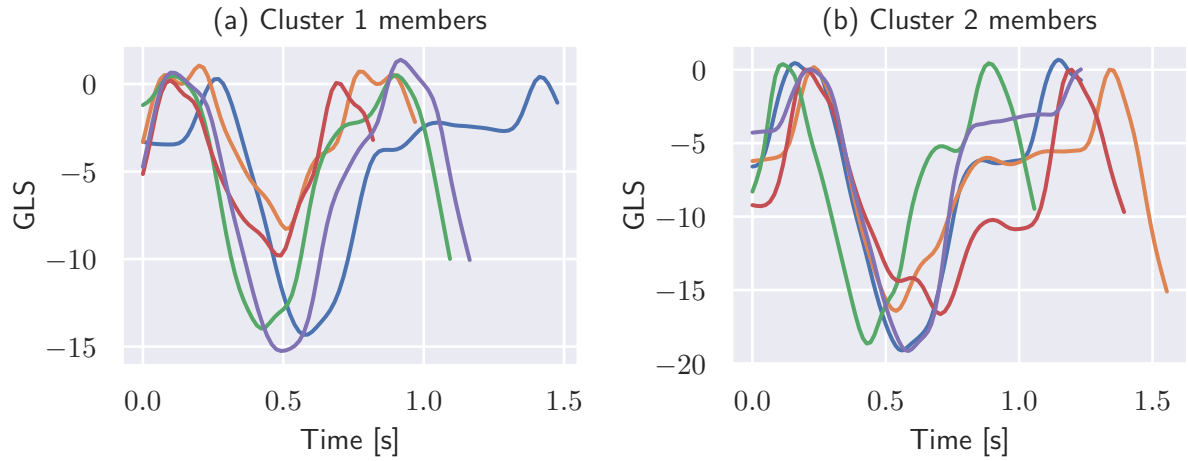


Figure 7.2: Here the curves of five random cluster members assigned by the *gls/2CH/regular/centroid/2* method. Each plot depicts the 2CH GLS curves for five random cluster members from the *gls/2CH/regular/centroid/2* method. (a) and (b) contain members from cluster 1 and 2 respectively. Only five curves are included to avoid making the plot to chaotic.

The majority of ARI scores are close to zero, but 17 methods evaluated at different numbers of cluster centers are able to acheive an ARI score above 0.20. As with DOR, the general trends for methods with a high ARI score is that they use data from a single view, use scaling or no preprocessing at all. From table 7.2 one can see that the top five methods only use the GLS curve from the 2CH view. In addition, one can also see that the two methods with the highest ARI (0.25) are the clustering methods evaluated at two cluster centers that perform best in terms of DOR as well. This means that there most likely are no methods evaluated at a number of cluster centers higher than two that will perform better than *gls/2CH/regular/centroid/2*, or *gls/2CH/scaled/centroid/2*. Figure 7.2 shows the 2CH GLS curves of five random cluster members from the *gls/2CH/regular/centroid/2* method. Although one cannot make any conclusive statements about what the general similarities between cluster members are, from the plots in figure 7.2 it seems like the curves of cluster 2 are smooth, while the curves of cluster 1 are more irregular in shape, which makes sense as this clustering algorithm uses a shape-based distance measure. Since *gls/2CH/regular/centroid/2* is one of two methods to acheive the highest DOR (11.72), accuracy (0.76), and ARI (0.25) it is chosen as the best of the TSC methods at identifying heart failure among patients. *gls/2CH/regular/centroid/2* is chosen over *gls/2CH/scaled/centroid/2* because it does not require preprocessing.

7.1.2 Peak-value Clustering

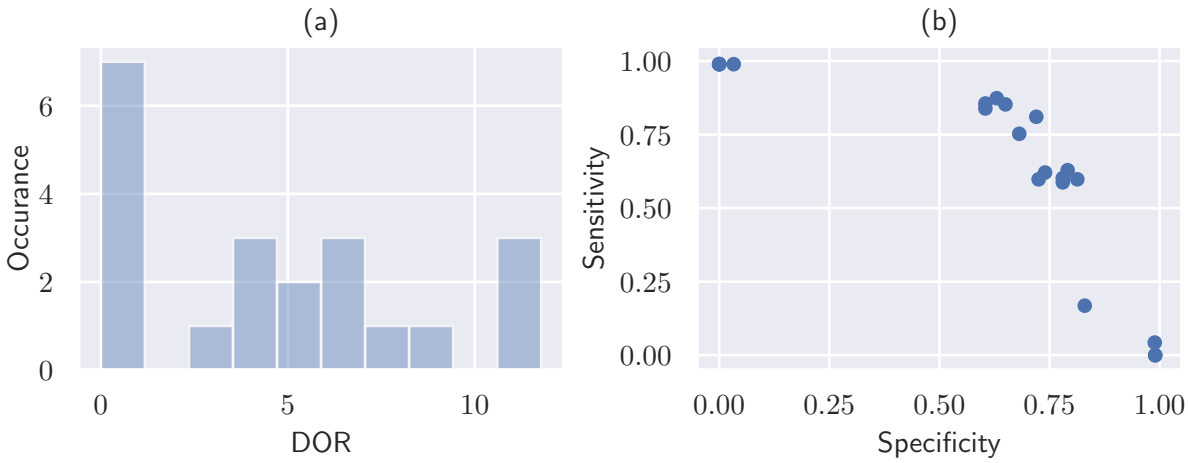


Figure 7.3: (a) Distribution plot of DOR of all PVC methods evaluated at two cluster centers when applied to classify heart failure. (b) Scatter plot of the same methods sensitivity, and specificity.

From figure 7.3a one can see that the majority of DOR scores are centered around zero, but there is a substantial number of methods that acheive a DOR score above 10. The scatterplot in figure 7.3b shows that there is also a great spread in sensitivity, and specificity. A few methods are spread along the edges of the plot acheiving a sensitivity or specificity score close to zero, but there are also methods that aceive sensitivity and specificity scores above 0.7. Common to the highest performing PVC methods is that they all use the dataset that is a combination of peak systolic GLS values and EF values. This can be confirmed from the complete table of results in the appendix 10.4. From table 7.3 one can see that *gls-EF/ward/2* is the PVC method that acheives the highest DOR of 11.59 when applied to classify heart failure. The *gls-EF/complete/2* method acheives the second highest DOR of 10.85, but its' specificity is nine points higher than *gls-EF/ward/2*, while its sensitivity is only six points lower, and it also has the highest accuracy of all the PVC methods applied to identify heart failure.

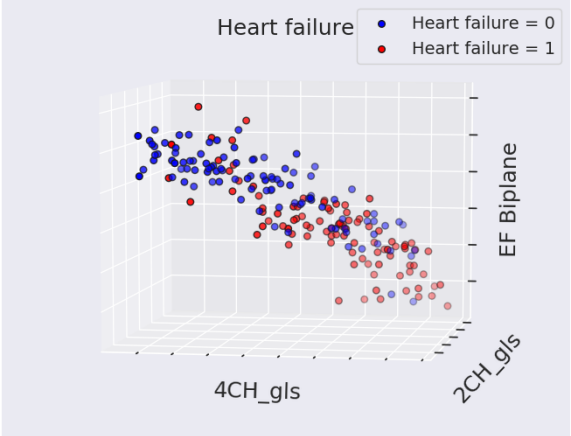
Dataset-Method	Accuracy	Sensitivity	Specificity	DOR
gls-EF/ward/2	0.75	0.87	0.63	11.59
gls-EF/complete/2	0.76	0.81	0.72	10.85
gls-EF/average/2	0.75	0.85	0.65	10.58
rls-EF/complete/2	0.73	0.86	0.60	8.89
gls-rls-EF/ward/2	0.72	0.84	0.60	7.80

Table 7.3: The accuracy, DOR, sensitivity and specicity scores of the five best performing two-cluster-center PVC methods in terms of DOR, at detecting heart failure. The **Dataset-Method** column indicates *Dataset used/Linkage criteria of method/Number of cluster centers.*

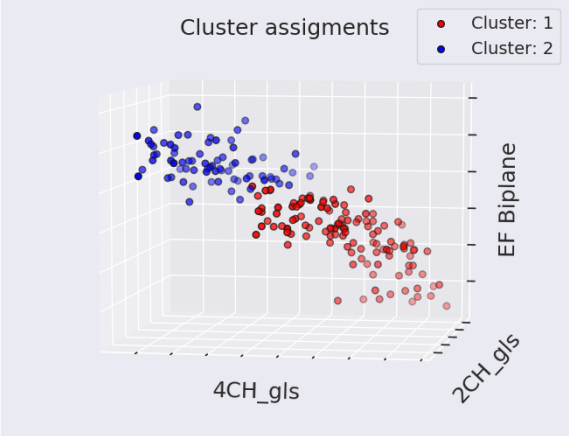
Dataset-Method	ARI
gls-EF/complete/2	0.27
gls-EF/ward/2	0.24
gls-EF/average/2	0.24
rls-EF/complete/2	0.21
gls-EF/complete/3	0.21

Table 7.4: The five highest ARI scores attained when applying PVC for detecting heart failure. The **Dataset-Method** column indicates *Dataset used/Linkage criteria of method/Number of cluster centers*.

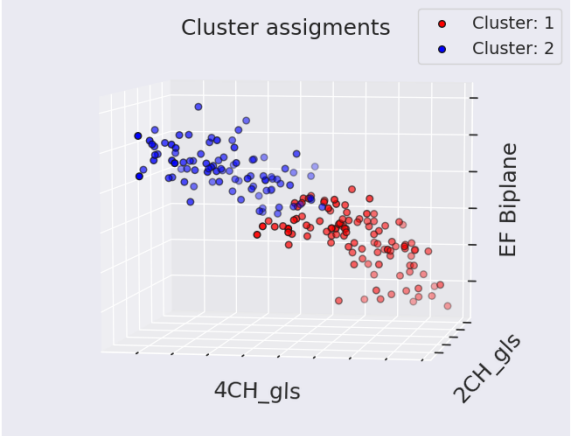
Many of the ARI of PVC methods for classifying heart failure are close to zero, but substantially more of the methods score above zero in ARI As with DOR, the methods that acheive the highest ARI scores use datasets that are combinations of strain curves and EF values. Table 7.4 shows that the three highest ARIs are attained by the same three methods that acheived the highest DORs. This means that there are most likely no methods evaluated at a higher number of cluster centers that will outperform *ward/2*, or *complete/2* at classifying heart failure. However, *complete/2* acheives the highest ARI, although it only acheives the second highest DOR. *complete/2* is chosen as the best performing PVC method when classifying heart failure, since it has the highest accuracy (76%), highest ARI (0.27), and second highest DOR (10.85). In figure 7.4 scatterplots patients are plotted with the dimensions: 4-chamber peak systolic GLS, 2-chamber peak systolic GLS and EF. The colors of the points correspond to wheather the patient has heart failure or not, and which cluster the points belong to. The plots are actually a lower dimensional projection of the GLS-EF peak-value dataset. This particular projection was chosen as it was found to be the projection where heart failure patients were as separable as possible. From plots 7.4b-d one can see that the clusters are fairly separable, heart failure on the other hand is not as easy to separate in these dimensions as can be seen in plot 7.4d. *Ward/2* and *complete/2* can in some sense be considered as binary classifiers where values under a certain threshold are categorized as heart failure. The *ward/2* method has the highest threshold for what is considered heart failure, and *complete/2* has the lowest, which explains their difference in sensitivity and specificity score. Since method *complete/2* acheives the highest accuracy (0.76), highest ARI (0.27) and second highest DOR (10.85) it is chosen as the best PVC method to identify heart failure among patients.



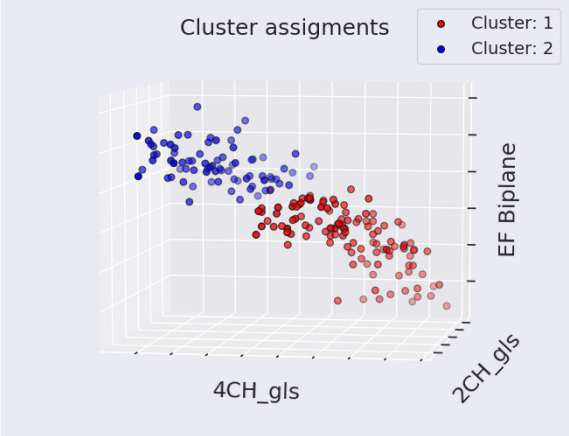
(a) Heart failure.



(b) *Ward/2* cluster assignments.



(c) *Complete/2* cluster assignments.



(d) *Average/2* cluster assignments.

Figure 7.4: Scatterplot of peak GLS values in each view. Colors in the of the different dots are given by heart failure diagnosis, and cluster assignments of ward/2, complete/2 and average/2 methods. Numbers are not included on the axes because the point of the figure is to illustrate the separability of clusters, and heart failure.

Dataset-Model	Accuracy	Sensitivity	Specificity	DOR
gls/4CH/upsampled	0.54	0.46	0.61	1.36
rls/APLAX/regular	0.53	0.48	0.58	1.30
rls/4CH/regular	0.52	0.36	0.68	1.20
gls/APLAX/downsampled	0.52	0.63	0.40	1.15
gls/2CH/downsampled	0.51	0.61	0.40	1.03

Table 7.5: The accuracy, DOR, sensitivity and specicity scores of the five best performing variations of the NN in terms of DOR, at detecting heart failure. The **Dataset-Model** column indicates *Dataset used/ View used/ Whether curve has been upsampled, downsampled or is regular.*

7.1.3 Deep Neural Network

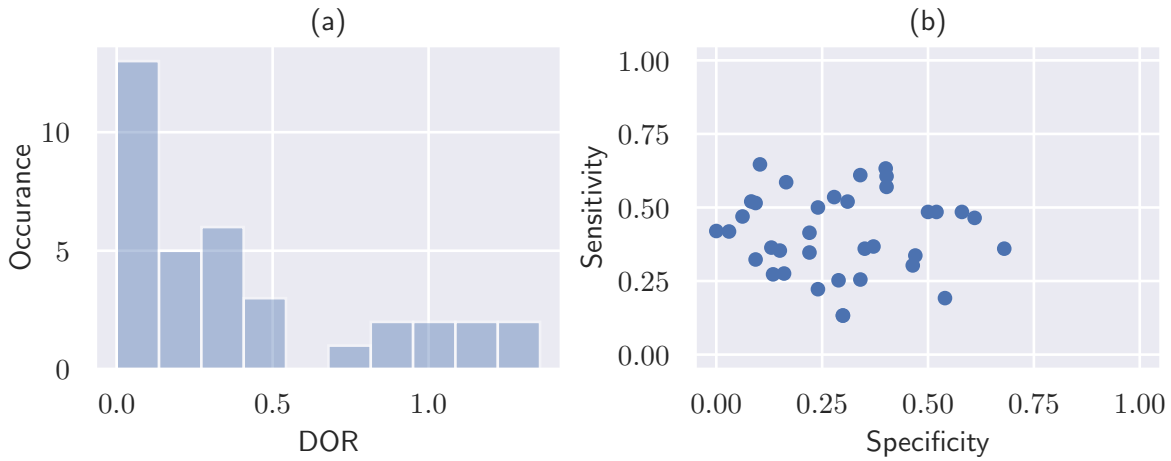


Figure 7.5: (a) Distribution plot of DOR of all NN models evaluated at two cluster centers when trained to predict heart failure. (b) Scatter plot of the same models sensitivity, and specificity.

From the distribution plot in figure 7.5a one can see that the most frequent DOR by NN models when training them to predict heart failure is zero . The highest DOR of 1.36 is attained by using only the GLS curve from the 4CH view as input, as can be seen from table 7.12. In the scatterplot in figure 7.5b one can see that sensitivity scores vary between 0.15 and 0.65, and the specificity scores vary between 0 and 0.68. The majority of the NN variations acheive a sensitivity, specificity and accuracy below 0.50. The accuracy of the model variations are also fairly low, 0.54 being the highest accuracy acheived. Since the heart failure dataset is fairly evenly distribution (recall figure 5.7) an accuracy of 0.54 is not much better than what could be acheived by randomly guessing the label. The 11 highest DORs attained by NN models trained to classify heart failure are acheived using only curves from single views as input, and only GLS, or RLS curves. *Gls/4CH/upsampled* will be considered the best model variation of the NNs at predicting heart failure since it acheives the highest accuracy and DOR .

Dataset-Model	Accuracy	Sensitivity	Specificity	DOR
gls-EF/Gaussian-Process	0.75	0.78	0.73	9.40
rls-EF/MLP	0.75	0.76	0.74	9.37
rls-EF/Linear-SVM	0.75	0.75	0.74	8.86
gls-EF/Ada-Boost	0.75	0.77	0.73	8.85
gls-EF/Naive-Bayes	0.75	0.76	0.74	8.79

Table 7.6: The accuracy, DOR, sensitivity and specicity scores of the five best performing PVSC in terms of DOR, at detecting heart failure. The **Dataset-Model** column indicates *Dataset used/The specific ML model used.*

7.1.4 Peak-value Supervised Classifiers

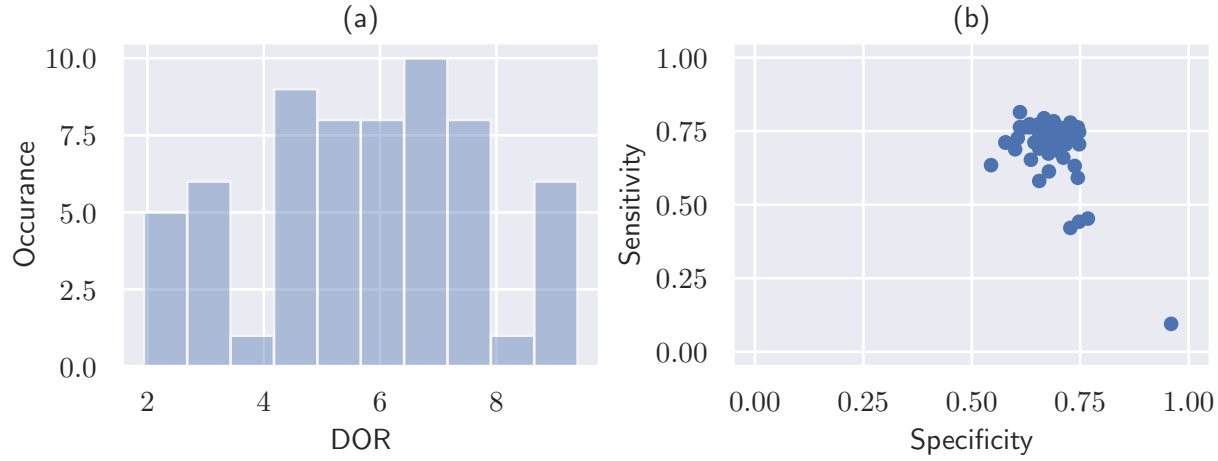


Figure 7.6: (a) Distribution plot of DOR of all PVSC models evaluated at two cluster centers when trained to predict heart failure. (b) Scatter plot of the same models sensitivity, and specificity.

From the distribution plot depicted in figure 7.6a one can see that the PVSC models overall acheive relatively high DORs, with a range of approximately two to nine. The scatterplot in figure 7.6b shows that the models are quite concentrated in terms of sensitivity and specificity scores. The majority of the models achieve sensitivity, and specificity scores in the ranges 0.6 to 0.75, with some outliers acheiving specificity below 0.5 and sensitivity above 0.75. What is even more concentrated are the accuracy scores of the models. As can be seen in table 7.6, the accuracy of top five PVSC models are all 0.75 As with PVC all the best performing PVSC models use a combination of EF and peak systolic strain values, and no specific ML model seems to outperform the others on all the datasets in term of DOR. The table also shows that the highest DOR of 9.4 is acheived by model *gls-EF/Gaissian-Process*. Although the DOR, sensitivity and specificity scores are very similar for the five best performing models *gls-EF/Gaussian-Process* is chosen as the PVSC model that performs best at predicting heart failure as it acheives the highest DOR.

Dataset-Model	Accuracy	Sensitivity	Specificity	DOR
TSC-gls/2CH/regular/centroid/2	0.76	0.87	0.64	11.72
PVC-gls-EF/complete/2	0.76	0.81	0.72	10.85
NN-gls/4CH/upsampled	0.54	0.46	0.61	1.36
PVSC-gls-EF/Gaussian-Process	0.75	0.78	0.73	9.40
Dataset-Model	TP	TN	FP	FN
TSC-gls/2CH/regular/centroid/2	86	62	35	13
PVC-gls-EF/complete/2	77	72	28	18
NN-gls/4CH/upsampled	46	61	39	53
PVSC-gls-EF/Gaussian-Process	74	72	27	21

Table 7.7: A table comparing the best contenders within each model group for predicting heart failure among patients. The top table comprare the models by their accuracy, sensitivity, specificity and DOR, and the bottom table shows the number of TPs, TNs, FPs and FNs that the different models attain.

7.1.5 Comparisons

With exeption of the NN, the models performance of the different models are very close in terms of DOR and accuracy. From table 7.7 one can see that the TSC method *gls/2CH/regular/centroid/2* achieves the highest sensitivity of all the models applied to predict heart failure, but it achieves the second lowest specificity of the four model groups. This can be confirmed by the fact that it attains 86 TPs, and 35 FPs. The PVSC model *gls-EF/Gaussian-Process* attains the most balanced score in terms of sensitivity and specificity, and the highest specificity score of all the model groups. However, the PVC model *gls-EF/complete/2* attains a higher accuracy, sensitivity and DOR than the PVSC model. One can also see that the PVC model attains more TP, the same number of TN, fewer FP and fewer FN than the PVSC model. It should also be noted that the PVC model and the PVSC model are using the same dataset which is a combination of peak systolic GLS values, and EF. To conclude this particular case study, the PVC model is picked as the best model at predicting heart failure among patients as it achieves the highest accuracy of the model groups, highest number of TN, and one of the most balanced combinations of sensitivity, and specificity.

Dataset-Method	Accuracy	Sensitivity	Specificity	DOR
gls/2CH/regular/centroid/2	0.74	0.71	0.93	33.47
gls/2CH/scaled/centroid/2	0.74	0.71	0.93	33.47
gls/2CH/scaled/average/2	0.73	0.69	0.93	30.71
gls/2CH/regular/average/2	0.73	0.69	0.93	30.71
gls/2CH/scaled/ward/2	0.71	0.67	0.93	27.49

Table 7.8: The accuracy, DOR, sensitivity and specicity scores of the five best performing two-cluster-center TSC methods in terms of DOR, at detecting patient diagnoses. The **Dataset-Method** column indicates *Dataset used/ View used/ Type of preprocessing used/ Linkage criteria of method/ Number of cluster centers.*

7.2 Case Study: Patient Diagnosis

7.2.1 Time-series Clustering

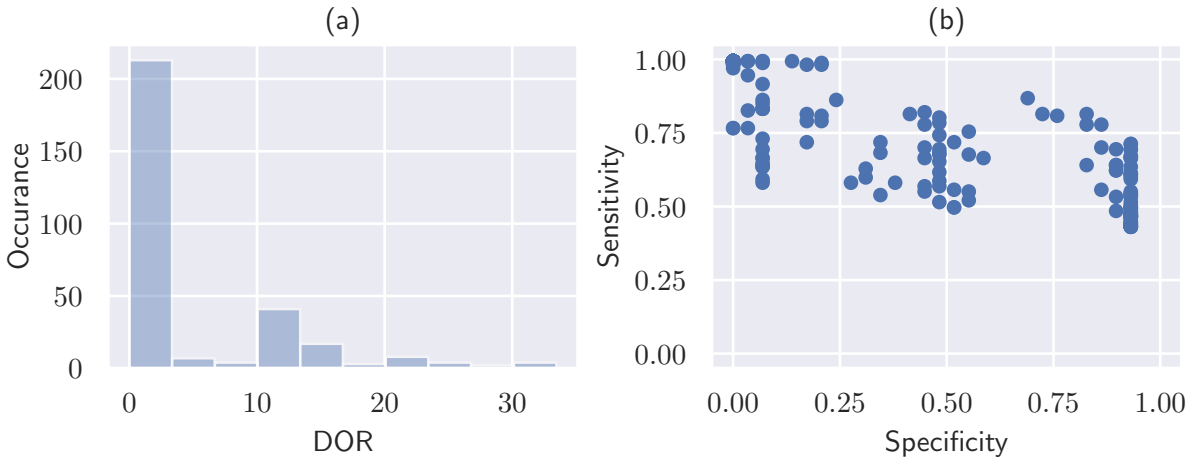


Figure 7.7: (a) Distribution plot of DOR of all TSC methods evaluated at two cluster centers when applied to classify patient diagnosis. (b) Scatter plot of the same methods sensitivity, and specificity.

From the distribution plot in figure 7.7a one can see that the majority of DORs are close to zero, but there are some methods that acheive a DOR above 30. In the scatter plot in figure 7.7b one can see that the specificity of the methods and range from 0.5 to 1, and the sensitivity scores range from 0 to 0.93. As with heart failure, the TSC methods that perform best in terms of DOR use data from a single view. The 2CH view, and GLS curves are the only view and curve that are used among the methods that achieve the five highest DORs. From the table of all the method results in the appendix 10.2 one can see that the highest performing method in terms of DOR to use a dataset other than GLS curves alone is *gls-rls/2CH/scaled/ward/2* and it achieves a DOR of 26.76. One can also note that the highest performing method in terms of DOR that uses a view other than only 2CH is *rls/all-views/normalized/weighted/2* which achieves a DOR of 25.56. The TSC methods that achieve the highest DOR scores all use no preprocessing, or scaling. From table 7.8 one can see that the TSC methods that acheive the highest DOR scores are *gls/2CH/regular/centroid/2*, and *gls/2CH/scaled/centroid/2* which are the same two methods that achieve the highest DORs in the heart failure case study.

Dataset-Method	ARI
gls-rls/4CH/regular/complete/2	0.36
gls/all-views/regular/weighted/2	0.34
gls/all-views/scaled/weighted/4	0.33
gls/all-views/scaled/weighted/3	0.33
gls/APLAX/regular/single/10	0.32

Table 7.9: The five highest ARI scores attained when applying TSC for detecting patient diagnoses. The **Dataset-Method** column indicates *Dataset used/ View used/ Linkage criteria of method/ Number of cluster centers.*

The majority of the ARI scorer for all the TSC methods evaluated at two to nine cluster centers are centered around zero. As with the TSC methods attaining the highest DORs the methods using no preprocessing or scaling acheive the highest ARI indices when used to identify patient diagnoses. In addition, the GLS curves are also most often part of the dataset for the TSC methods receiving the highest ARI when used to identify patient diagnoses. From table 7.9 one can see that the TSC methods receiving the five highest ARI scores, are not among the TSC methods that receive the highest DOR scores. The TSC method *gls-rls/4CH/regular/complete/2* attains the highest ARI score when applied to identify patient diagnoses, and achieves an accuracy of 0.84, a sensitivity of 0.87 a specificity of 0.69 and a DOR 14.65. The TSC method *gls/all-views/regular/weighted/2* achieves the second highest ARI when applied to identify patient diagnoses, and achieves an accuracy of 0.82, a sensitivity of 0.81 a specificity of 0.83 and a DOR 21.06. What should also be noted is that the TSC methods achieving the two highest ARIs when applied to identify patient diagnoses are methods evaluated at two cluster centers, which means that none of the TSC methods evaluated at cluster centers between three and nine can perform better than the ones evaluated at two cluster centers. It may seem strange that the ordered lists of DORs, and ARIs are so different. The reason for this is not because DOR inherently values sensitivity higher than specificity, but stems from how the DOR is defined. Recall that $\text{DOR} = (\text{TP} \times \text{TN})/(\text{FP} \times \text{FN})$, since the patient diagnoses dataset is skewed in favour of positives TP has the potential of being as high as 170 while TN can be as high as 30. Therefore the DOR will be higher for methods with a high sensitivity than for methods with an equally high sensitivity. In figure 7.8 curves of five random cluster members assigned by the *gls/all-views/regular/weighted/2* method are plotted. As with the observations made with regard to figure 7.2 it is not possible to make any conclusive statements as to what the similarities are based on such a small sample size. However, based on the small sample size in 7.8 it seems as though the curves in cluster 2 (column (b)) are smoother in shape, than the curves in cluster 1 (column (a)). The TSC method that is chosen as the best method for identifying patient diagnoses is *gls/all-views/regular/weighted/2*, because it achieves the second highest ARI, and because it's sensitivity and specificity are more balanced than the method attaining the highest ARI and the methods that achieve higher DORs.

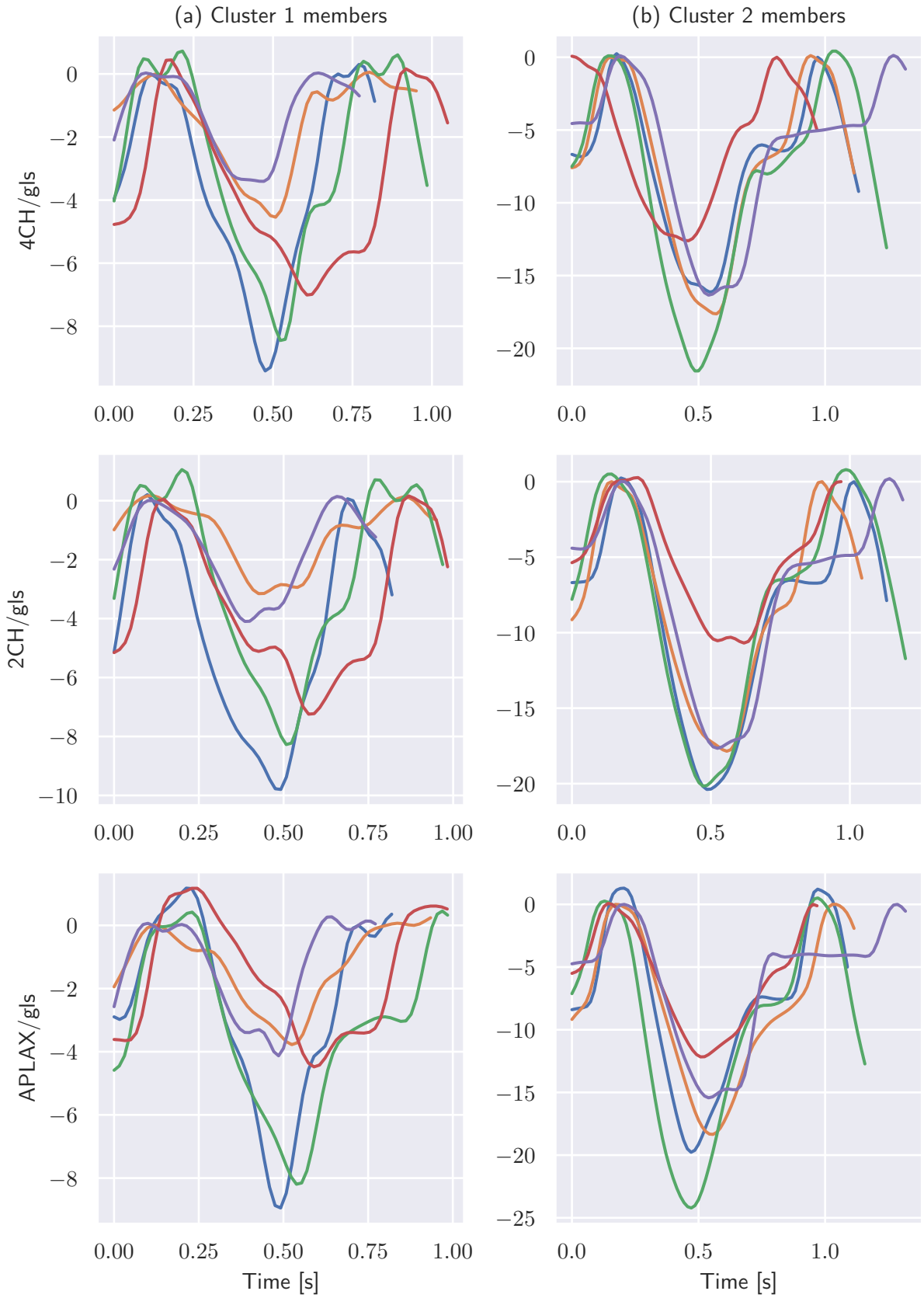


Figure 7.8: Here the curves of five random cluster members assigned by the *gls/all-views/regular/weighted/2* method are plotted. Each row represents one of the seven possible strain curves in the 4CH view. Coloumn (a) and (b) represent cluster 1 and 2 respectively. To make it easier to visually separate the curves, only five random members from cluster 1 and 2 are included in the figure.

Dataset-Method	Accuracy	Sensitivity	Specificity	DOR
gls-EF/ward/2	0.76	0.72	0.94	39.33
rls-EF/complete/2	0.77	0.74	0.93	37.61
gls-rls-EF/ward/2	0.76	0.72	0.93	35.16
gls-EF/average/2	0.74	0.70	0.94	34.90
gls-EF/complete/2	0.68	0.63	0.94	25.75

Table 7.10: The accuracy, DOR, sensitivity and specicity scores of the five best performing two-cluster-center PVC methods in terms of DOR, at detecting patient diagnoses. The **Dataset-Method** column indicates *Dataset used/Linkage criteria of method/Number of cluster centers*.

7.2.2 Peak-value Clustering

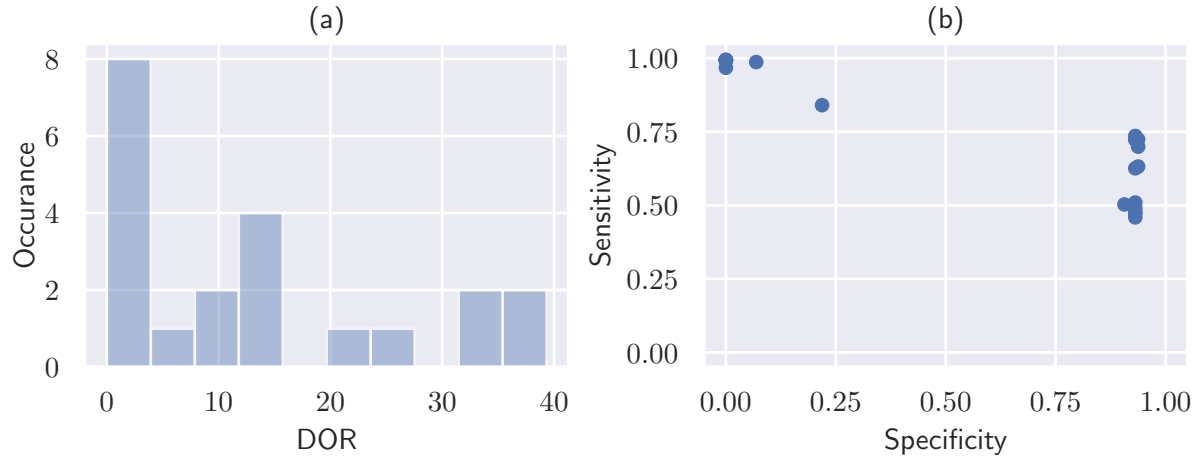


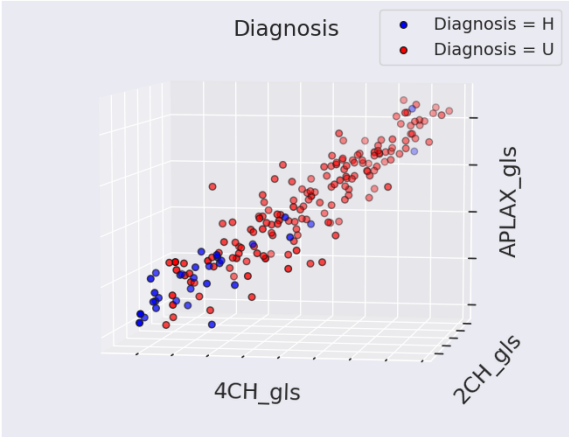
Figure 7.9: (a) Distribution plot of DOR of all PVC methods evaluated at two cluster centers when applied to classify patient diagnosis. (b) Scatter plot of the same methods sensitivity, and specificity.

From the distribution plot in figure 7.9a one can see that the majority of the PVC methods get DORs close to zero, but there are a few methods that attain DORs above 30, and close to 40. From the scatter plot in 7.9b one can see that almost all the sensitivity scores are above 0.5, while the specificity scores are concentrated in the areas 0 to 0.25 and 0.95. As with the heart failure case study the PVC methods that perform high in terms of DOR use a dataset that is a combination of peak systolic strain values and EF. From table 7.10 one can see that *gls-EF/ward/2* and *rls-EF/complete/2* are the two top performers in terms of DOR. *gls-EF/ward/2* achieves a slightly higher specificity score, where as *rls-EF/complete/2* attains a slightly higher specificity score.
The majority of the ARI scores of PVC methods applied to identify patient diagnoses are centered around zero, but as one can see from table 7.11 there are a few methods that acieve an ARI above 0.2 close to 0.3. For a change, the PVC methods that perform best in terms of ARI, are neither methods evaluated at two cluster centers, or methods that are applied on a combination of peak systolic strain values and EF. In contrast to the heart failure case study, the PVC methods that achieve the highest ARIs, when applied to identify patient diagnoses, are not the same methods that achieve the highest DORs. The two PVC methods that achieve the highest ARIs are the *gls/average* method evaluated at 6 and 7 cluster centers respectively.

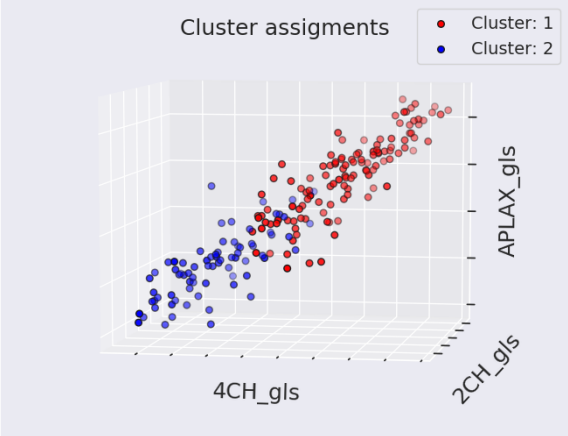
Dataset-Method	ARI
gls/average/6	0.29
gls/average/7	0.29
gls-rls/complete/3	0.28
rls-EF/complete/2	0.26
gls-EF/ward/2	0.25

Table 7.11: The five highest ARI scores attained when applying PVC for detecting patient diagnoses. The **Dataset-Method** column indicates *Dataset used/Linkage criteria of method/Number of cluster centers.*

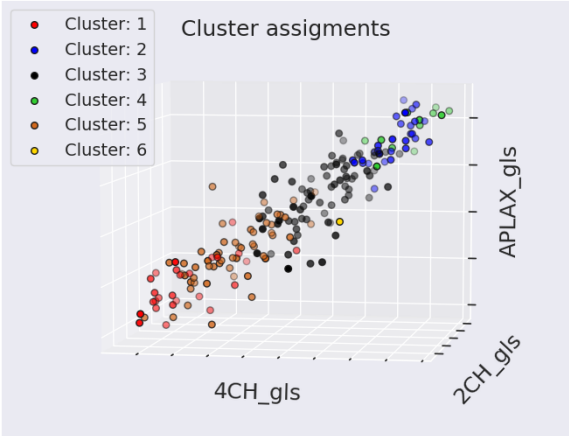
To get a better idea of why *gls/average/6* and *gls/average/7* attain the ARIs they do, scatter plots of these two methods, and *gls-EF/ward/2* have been given in figure 7.4. A scatter plot of the target variable patient diagnosis is also given for comparison. The dimensions used are peak systolic GLS in all three views as these are the dimensions that are common to all three methods. From the scatter plot in plot 7.10a one can see that the healthy patients are in the minority, and are concentrated in the corner with low peak systolic GLS values in the 4CH, 2CH and APLAX views. There are also some healthy patients with low-medium peak systolic GLS values, and very few healthy patients with high peak systolic GLS values. From plot 7.10b one can see that *gls-EF/ward/2* is able to isolate the concentration of healthy patients with low peak systolic GLS, but at the cost of many FNs. * In plot 7.10c and 7.10d one can see that cluster 1 of method *gls/average/6*, and cluster 2 of method *gls/average/7* capture the healthy patients with low peak systolic GLS, but are unable of capturing the healthy patients with medium to high values. If one combines clusters 1 and 5 of *gls/average/6*, and lets them represent healthy patients, and let the remaining clusters represent unhealthy patients the method attains an accuracy of 0.74, a sensitivity of 0.70, a specificity of 0.94 and a DOR of 34.90. If one combines clusters 2 and 5 of *gls/average/7*, and lets them represent healthy patients, and let the remaining clusters represent unhealthy patients this method attains an accuracy of 0.74, a sensitivity of 0.70, a specificity of 0.94 and a DOR of 35.94. While the performance of the revised *gls/average/6* and *gls/average/6* methods are good, they are still not as good as the performance of the top three performers in terms of DOR, which attain higher accuracy, sensitivity and DORs. Therefore, *rls-EF/complete/2* is chosen as the best of the PVC methods at identifying patient diagnosis, as it achieves the second highest DOR, and a more balanced sensitivity/specificity than *gls-EF/ward/2* that attains the highest DOR score.



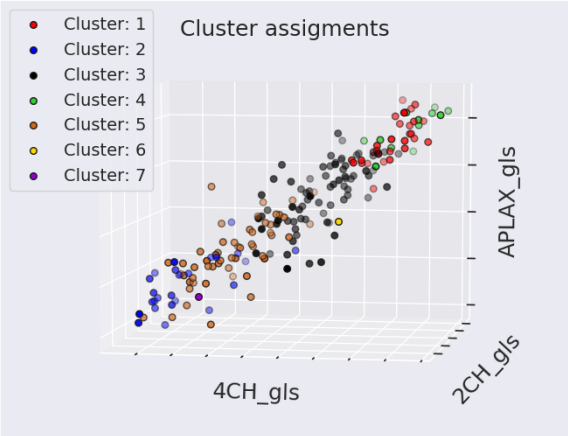
(a) Patient Diagnosis. **H** stands for **Healthy**, and **U** stands for **Unhealthy**



(b) *GLS-EF Ward/2* cluster assignments.



(c) *GLS Average/6* cluster assignments.



(d) *GLS Average/7* cluster assignments.

Figure 7.10: Scatterplot of peak GLS values in each view. Colors in the of the different dots are given by heart failure diagnosis, and cluster assignments of *gls-EF/ward/2*, *average/6* and *average/7* methods. Numbers are not included on the axes because the point of the figure is to illustrate the separability of clusters, and patient diagnosis.

Dataset-Model	Accuracy	Sensitivity	Specificity	DOR
all-strain/4CH/upsampled	0.83	0.99	0.00	0.00
all-strain/2CH/regular	0.85	1.00	0.00	NaN
gls/2CH/regular	0.85	1.00	0.00	NaN
rls/2CH/regular	0.85	1.00	0.00	NaN
all-strain/2CH/downsampled	0.85	1.00	0.00	NaN

Table 7.12: The accuracy, DOR, sensitivity and specicity scores of the five best performing variations of the NN in terms of DOR, when trained to predict patient diagnoses. The **Dataset-Model** column indicates *Dataset used/ View used/ Whether curve has been upsampled, downsampled or is regular.*

7.2.3 Deep Neural Network

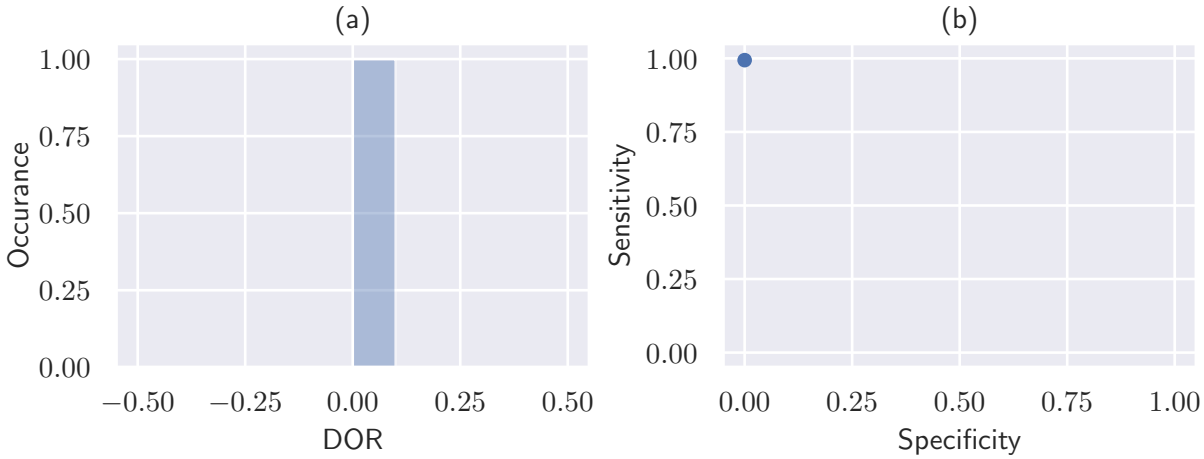


Figure 7.11: (a) Distribution plot of DOR of all NN models when trained to classify patient diagnosis. (b) Scatter plot of the same methods sensitivity, and specificity.

From the distribution plot in figure 7.11 one can see that the collective performance of the different variations of the NN trained to predict patient diagnosis is terrible. The DOR of all the methods are either zero because the number of TNs attained are zero, or not defined because the number of FNs are zero. The sensitivities are all 1, or close to 1, and the specificities are all 0. It is evident that the NNs are not able to generalize the traits of the healthy patients from such a small dataset. The NN models are will therefore not be discussed further with relation to prediction of patient diagnosis, and are not included in the comparison of the four model groups.

Dataset-Model	Accuracy	Sensitivity	Specificity	DOR
gls-rls-EF/Ada-Boost	0.95	0.97	0.79	138.42
gls-rls/KNN	0.93	0.95	0.82	84.53
rls-EF/Extra-Trees	0.93	0.96	0.75	76.50
gls-rls-EF/Extra-Trees	0.93	0.97	0.71	75.00
gls-rls/Extra-Trees	0.93	0.97	0.71	75.00

Table 7.13: The accuracy, DOR, sensitivity and specicity scores of the five best performing PVSC models in terms of DOR, when trained to predict patient diagnosis. The **Dataset-Model** column indicates *Dataset used/Specific machine learning model used.*

7.2.4 Peak-value Classifiers

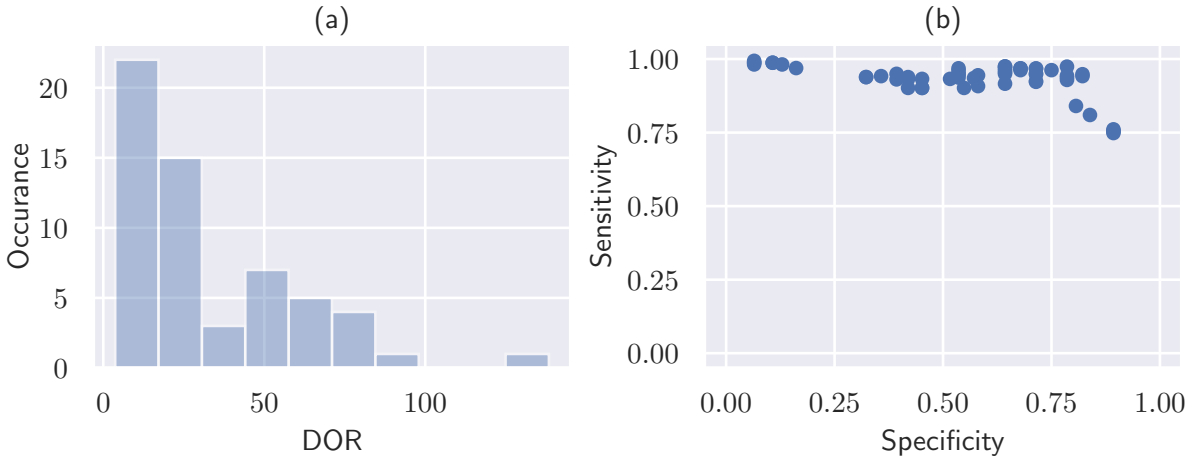


Figure 7.12: (a) Distribution plot of DOR of all PVSC models when trained to classify patient diagnosis. (b) Scatter plot of the same methods sensitivity, and specificity.

From the distribution plot in figure **??** one can see that the majority of the DOR scores are close to zero, but there are couple of methods that attain DORs close to 100, and one method that attains a DOR close to 150. From the scatter plot in figure **??** one can see that the sensitivity ranges from 0.75 to 1, and the specificity ranges from close to zero to approximately 0.95. Among the top five PVSC models in terms of DOR are many different combinations of models, and datasets. Three of the five highest DOR scores are attained by Extra-Trees models, and the top two scores are attained by KNN and Ada Boost classifiers. *gls-rls-EF/Ada-Boost* and *gls-rls/KNN* are the two top PVSC performers with regard to DOR. *gls-rls-EF/Ada-Boost* achieves the highest sensitivity of the two by two points, and *gls-rls/KNN* achieves the highest specificity of the two by three points. Since sensitivity and specificity is weighted equally in this study *gls-rls/KNN* is chosen as the best of the PVSC models trained to identify patient diagnoses.

Dataset-Model	Accuracy	Sensitivity	Specificity	DOR
TSC-gls/all-views/regular/weighted/2	0.82	0.81	0.83	21.06
PVC-rls-EF/complete/2	0.77	0.74	0.93	37.61
PVSC-gls-rls/KNN	0.93	0.95	0.82	84.53
Dataset-Model	TP	TN	FP	FN
TSC-gls/all-views/regular/weighted/2	136	24	5	31
PVC-rls-EF/complete/2	117	27	2	42
PVSC-gls-rls/KNN	147	23	5	4

Table 7.14: A table comparing the best contenders within each model group for predicting patient diagnoses. The top table comprare the models by their accuracy, sensitivity, specificity and DOR, and the bottom table shows the number of TPs, TNs, FPs and FNs that the different models attain on their respective datasets.

7.2.5 Comparisons

From the top table in 7.14 one can see that there is a significant difference in performance between the three models included for comparison. The TSC model *gls/all-views/regular/weighted/2* attains the second highest accuracy, sensitivity and specificity of the three models, but also attains the lowest DOR. The TSC model can also be said to attain the most balanced scores in terms of sensitivity and specificity. The PVC model *rls-EF/complete/2* attains the highest specificity, second highest DOR, but lowest sensitivity and accuracy of the three models. The PVSC model *gls-rls/KNN* attains the highest accuracy, sensitivity and DOR of all the models, but it also achieves the lowest specificity of all the models. However, since the PVSC model is so close to the TSC model in terms of specificity, and is so much better than the other two models in all other metrics, it is chosen as the best model of identifying patient diagnoses. This can be confirmed from the bottom table in 7.14, where one can see that the PVSC model only gets one TN less than the TSC model, but attains 11 more TP.

7.3 Case Study: Segment Indication

7.3.1 Time-series Clustering

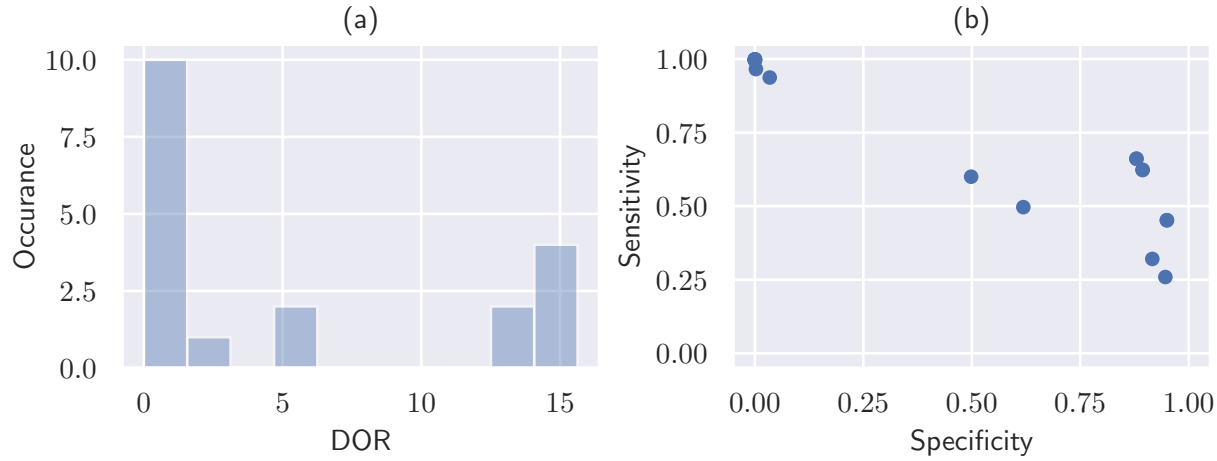


Figure 7.13: Distribution of DOR, sensitivity and specificity for the different TSC methods when classifying left ventrice segment indication.

Dataset-Method	Accuracy	Sensitivity	Specificity	DOR
regular/weighted/2	0.69	0.45	0.95	15.63
scaled/weighted/2	0.69	0.45	0.95	15.63
regular/ward/2	0.77	0.66	0.88	14.26
scaled/ward/2	0.77	0.66	0.88	14.26
regular/complete/2	0.75	0.62	0.89	13.92

Table 7.15: The accuracy, DOR, sensitivity and specicity scores of the five best performing two-cluster-center TSC methods in terms of DOR, at detecting segment indication. The **Dataset-Method** column indicates *Type of preprocessing used/Linkage criteria of method/Number of cluster centers.*

From the distribution plot in figure 7.13a one can see that the majority of the DORs are close to zero, but a few methods are able to achieve DORs above 12, and some methods attain a DOR close to 15 when applied to identify segment indication. From the scatter plot in figure 7.13b one can see that the sensitivity of the TSC methods range from 0.25 to 1, and the specificity of the TSC methods range from 0 to approximately 1. The spread in both sensitivity and specificity is quite large, and there are very few methods that are able to a attain a high sensitivity while at the same time attaining a high specificity, and vice versa. Common to the high performing TSC methods in terms of DOR is that they all use either no preprocessing at all, or scaling. *z-norm/complete/2* is the seventh best TSC method in terms of DOR, and attains a DOR of 5.92 when applied to identify segment indication. *norm/ward/2* is the ninth best methods in terms of DOR, and attains a DOR of 1.56, when applied to identify segment indication. This can be comfirmed from table 10.3 The two TSC methods attaining the highest DORs *regular/weighted/2*, and *scaled/weighted/2* differ only in type of preprocessing used. From table 7.15 and table 10.3 one can see that the two methods attain the same scores in all metrics, this is because they yield the exact same cluster assignments to the individual segment strain curves.

Dataset-Method	ARI
scaled/centroid/5	0.286
regular/centroid/5	0.286
regular/ward/2	0.284
scaled/ward/2	0.284
scaled/centroid/6	0.271

Table 7.16: The five highest ARI scores attained when applying TSC for detecting segmend indication. The **Dataset-Method** column indicates *Type of preprocessing used/Linkage criteria of method/Number of cluster centers.*

The same goes for the next two TSC methods in line *regular/ward/2 scaled/ward/2*, these two methods are also the methods that attain the highest accarcy of all the TSC methods. Of the two TSC methods *regular/weighted/2*, and *regular/ward/2* the latter is preferred for predicting segment indication because *regular/ward/2* has a more persistent performance in both sensitivity and specificity, where as *regular/weighted/2* has a high specificity, but a very low sensitivity.
The majority of the ARIs of TSC methods applied to identify segment indication, but as one can see from table 7.16 some methods are able to attain ARIs above 25. As with the other case studies, the TSC methods that attain the highest ARIs are methods that use either no preprocessing at all or scaling. Puzzlingly enough the top two TSC methods for classifying segment indication in terms of ARI, are methods evaluated at five cluster centers, not two. TSC methods *scaled/centroid/5*, and *regular/centroid/5* differ only in type of preprocessing used, and they yield the exact same cluster assignments, and evaluations scores. The next two methods in order of ARI *regular/ward/2*, and *scaled/ward/2* are familiar from the list of TSC methods attaining the highest DORs when applied to identify segment indication. From table 7.16 one can also see that the difference in ARI between *regular/centroid/5*, and *regular/ward/2* is only 0.002 Since the *regular/ward/2* method will be considered the best of the TSC methods at classifying segment indication. It attains the third highest ARI of all the TSC methods applied to identify segment indication, and is the preferred method among the TSC methods evaluated at two cluster centers.

7.3.2 Deep Neural Network

Method	Accuracy	Sensitivity	Specificity	DOR
regular	0.74	0.80	0.68	8.65
downsampled	0.74	0.74	0.75	8.38
upsampled	0.65	0.55	0.73	3.36

Table 7.17: Evaluation metrics of the NN for classifying the binary indication of individual segments in the left ventricle.

Of the three variations of the NN model, the one that uses no resampling, and the one that downsamples all signals to the lowest sample rate achieve relatively similar DOR scores. The variation that upsamples the sample rate of all the curves to the highest sample rate performs significantly worse than the other two in terms of DOR and sensitivity. Of the three variations the model that uses downsampling is the preferred model of the three since its sensitivity and specificity are more balanced than the method that uses no resampling, and accuracy is higher than the method that uses upsampling.

7.3.3 Comparisons

From table 7.18 one can see that the performances of the NN, and TSC models are quite close in terms of accuracy, but differ significantly in the other metrics. The TSC model *regular/ward/2* attains a higher accuracy, specificity and DOR than the NN model *downsampled*. This can also be confirmed by the fact that the TSC model attains more TN, and fewer FP than the NN model. The NN model attains the highest sensitivity, which can be confirmed by the fact that it attains more TP and fewer FN than the TSC model. The NN model is also the model that attains the most balanced scores of sensitivity and specificity. Therefore the NN model is chosen as the best performer at predicting the segment indication.

Dataset-Model	Accuracy	Sensitivity	Specificity	DOR
TSC-regular/ward/2	0.76	0.64	0.88	13.15
NN-downsampled	0.74	0.74	0.75	8.38
Dataset-Model	TP	TN	FP	FN
TSC-regular/ward/2	1202	1491	204	616
NN-downsampled	1255	1390	473	440

Table 7.18: A table comparing the best contenders within each model group for predicting segment indication. The top table comprare the models by their accuracy, sensitivity, specificity and DOR, and the bottom table shows the number of TPs, TNs, FPs and FNs that the different models attain.

Chapter Summary

In the heart failure case study the PVC model was found to be the best performer, by a narrow margin. The TSC, and PVSC models also performed well, but the NN did not. In fact, the performance of the NN was not much better than what could be achieved by randomly guessing the binary label with equal probability of choosing one or zero. The PVC model that performed best at identifying heart failure among patients is *gls-EF/complete/2*, and it attains an accuracy of 0.76, sensitivity of 0.81, specificity of 0.72 and DOR of 10.85

In the patient diagnosis case study the PVSC model is regarded as the top performer. Here too, it was a close call between the PVSC, PVC and TSC models. The patient diagnosis dataset was skewed as there were 170 patients with a heart disease, and only 30 healthy patients. For this reason it is probable that the NN was unable to generalize the feature of the healthy patients, because almost all the variations of the NN ended up always making the prediction that the patient was diseased yielding a score of 0 in specificity. The PVSC model that performed best at predicting patient diagnosis is *gls-rls/KNN*, and it attains an accuracy of 0.93, sensitivity of 0.95, specificity of 0.82 and DOR of 84.53

In the segment indication case study only the TSC and NN models were compared, and for a change of pace it was only the NN that was chosen as the best performer. The TSC model did not perform much worse, in fact it performed better than the NN in many respects. The key reason for why the NN was preferred was because it had a more balanced sensitivity, and specificity scores than the TSC model. The NN model that performed best at predicting segment indication is *downsampled*, and it attains an accuracy of 0.74, sensitivity of 0.74, specificity of 0.72 and DOR of 8.38

Chapter 8

Discussion

In the results chapter, the performance results were presented in the order of the different target variables that were explored. In the discussion chapter a different approach is taken, and the each model will be discussed individually based on their performance in the case studies.

8.1 Time-series Clustering

- Discuss why methods using data from single view performed better.
- Discuss why methods using data using "regular" or "scaled" performed better than methods using normalization or z-normalization.

8.2 Peak-value Clustering

- Comment on the fact that TSC generally performs better than PVC.
- Note that the PVC methods that use combination of peak systolic strain values and EF performs significantly better, future work could be to combine EF values with strain curves as well.
-
-
-

8.3 Neural Networks

-

Chapter 9

Conclusion

This is the conclusion.

9.1 Future Work

This is the future work section.

Chapter 10

Appendix

This is the appendix

10.1 Raw model results

10.1.1 Time-series Clustering

Table 10.1: Classification results of applying TSC to identify heart failure among patients. The results are sorted in descending order of DOR, although DOR is not included.

Dataset-Method	TP	TN	FP	FN
gls/2CH/regular/centroid/2	86	62	35	13
gls/2CH/scaled/centroid/2	86	62	35	13
gls/2CH/regular/average/2	84	63	34	15
gls/2CH/scaled/average/2	84	63	34	15
gls-rls/2CH/scaled/ward/2	81	65	32	18
rls/APLAX/scaled/weighted/2	90	45	52	9
gls-rls/APLAX/regular/median/2	26	93	4	73
gls-rls/APLAX/scaled/weighted/2	90	43	54	9
rls/APLAX/scaled/average/2	82	60	37	17
gls/2CH/regular/ward/2	80	63	34	19
gls/2CH/scaled/ward/2	80	63	34	19
gls-rls/2CH/scaled/complete/2	74	70	27	25
gls-rls/4CH/regular/weighted/2	98	7	90	1
rls/2CH/scaled/ward/2	77	66	31	22
gls/2CH/regular/complete/2	75	68	29	24
gls/2CH/scaled/complete/2	75	68	29	24
gls/4CH/scaled/centroid/2	77	64	33	22
gls/4CH/regular/centroid/2	77	64	33	22
rls/all-views/regular/complete/2	86	49	48	13
gls/all-views/regular/weighted/2	88	44	53	11
gls/all-views/regular/centroid/2	74	67	30	25
gls-rls/2CH/regular/complete/2	73	68	29	26
gls-rls/2CH/regular/ward/2	78	62	35	21
gls-rls/APLAX/scaled/average/2	81	57	40	18
gls/all-views/scaled/average/2	73	67	30	26
gls/all-views/regular/median/2	21	93	4	78
gls-rls/4CH/regular/complete/2	91	34	63	8

gls/4CH/regular/complete/2	74	64	33	25
gls/4CH/scaled/complete/2	74	64	33	25
gls/all-views/scaled/ward/2	67	71	26	32
gls/all-views/regular/ward/2	67	71	26	32
gls-rls/4CH/scaled/weighted/2	85	47	50	14
gls/all-views/scaled/complete/2	66	71	26	33
rls/2CH/regular/complete/2	67	70	27	32
gls/all-views/regular/average/2	62	74	23	37
gls/all-views/regular/complete/2	62	74	23	37
rls/all-views/scaled/ward/2	59	76	21	40
gls-rls/4CH/scaled/average/2	60	75	22	39
gls-rls/all-views/regular/complete/2	60	75	22	39
gls-rls/all-views/scaled/weighted/2	60	75	22	39
gls/APLAX/regular/ward/2	65	71	26	34
gls/APLAX/regular/median/2	65	71	26	34
gls-rls/all-views/regular/ward/2	61	74	23	38
rls/all-views/scaled/weighted/2	58	76	21	41
gls-rls/APLAX/scaled/centroid/2	58	76	21	41
gls-rls/all-views/regular/centroid/2	62	73	24	37
gls/APLAX/regular/average/2	63	72	25	36
rls/APLAX/scaled/ward/2	59	75	22	40
rls/all-views/scaled/complete/2	59	75	22	40
gls-rls/APLAX/scaled/complete/2	41	85	12	58
gls/APLAX/regular/centroid/2	65	70	27	34
gls-rls/all-views/scaled/average/2	60	74	23	39
gls/APLAX/regular/complete/2	47	82	15	52
gls/all-views/scaled/centroid/2	61	73	24	38
gls-rls/all-views/scaled/centroid/2	61	73	24	38
gls/4CH/regular/median/2	24	91	6	75
gls/4CH/regular/weighted/2	24	91	6	75
gls/4CH/scaled/weighted/2	24	91	6	75
gls/4CH/scaled/median/2	24	91	6	75
gls-rls/APLAX/regular/average/2	58	75	22	41
rls/APLAX/regular/ward/2	58	75	22	41
gls-rls/all-views/regular/average/2	58	75	22	41
rls/APLAX/regular/weighted/2	46	82	15	53
gls/all-views/scaled/weighted/2	13	94	3	86
gls-rls/4CH/scaled/ward/2	56	76	21	43
rls/4CH/scaled/average/2	56	76	21	43
rls/all-views/scaled/average/2	54	77	20	45
gls-rls/APLAX/scaled/ward/2	54	77	20	45
rls/all-views/regular/average/2	54	77	20	45
gls-rls/all-views/scaled/complete/2	54	77	20	45
gls-rls/4CH/scaled/complete/2	54	77	20	45
rls/APLAX/scaled/complete/2	58	74	23	41
gls-rls/APLAX/regular/ward/2	55	76	21	44
gls-rls/all-views/scaled/ward/2	55	76	21	44
rls/4CH/regular/complete/2	64	69	28	35
gls/APLAX/regular/weighted/2	36	86	11	63
gls/2CH/scaled/median/2	57	74	23	42
gls/2CH/scaled/weighted/2	57	74	23	42

gls/2CH/regular/weighted/2	57	74	23	42
gls/2CH/regular/median/2	57	74	23	42
gls-rls/APLAX/regular/centroid/2	51	78	19	48
gls-rls/4CH/regular/ward/2	54	76	21	45
gls-rls/4CH/regular/average/2	54	76	21	45
rls/APLAX/regular/complete/2	54	76	21	45
rls/4CH/scaled/ward/2	52	77	20	47
rls/2CH/normalized/median/2	30	88	9	69
rls/all-views/normalized/weighted/2	98	4	93	1
rls/APLAX/normalized/median/2	98	4	93	1
rls/2CH/scaled/complete/2	60	71	26	39
gls/4CH/scaled/ward/2	43	82	15	56
gls/4CH/regular/ward/2	43	82	15	56
gls-rls/APLAX/regular/complete/2	73	58	39	26
rls/all-views/regular/ward/2	54	75	22	45
gls-rls/all-views/regular/weighted/2	46	80	17	53
gls/all-views/scaled/median/2	65	65	32	34
rls/4CH/scaled/complete/2	58	71	26	41
rls/4CH/regular/ward/2	52	75	22	47
rls/2CH/regular/ward/2	45	79	18	54
gls-rls/APLAX/regular/weighted/2	29	86	11	70
rls/4CH/regular/weighted/2	97	6	91	2
gls-rls/all-views/normalized/ward/2	27	85	12	72
rls/all-views/normalized/complete/2	35	80	17	64
gls-rls/4CH/normalized/ward/2	53	65	32	46
gls-rls/4CH/z-normalized/ward/2	60	58	39	39
gls-rls/2CH/z-normalized/ward/2	78	36	61	21
gls-rls/4CH/scaled/median/2	96	6	91	3
rls/4CH/z-normalized/ward/2	60	56	41	39
rls/2CH/z-normalized/weighted/2	98	2	95	1
rls/all-views/scaled/median/2	98	2	95	1
gls-rls/2CH/z-normalized/complete/2	98	2	95	1
rls/all-views/normalized/ward/2	58	56	41	41
gls/2CH/z-normalized/ward/2	70	42	55	29
rls/all-views/z-normalized/ward/2	50	61	36	49
gls-rls/APLAX/normalized/ward/2	25	81	16	74
gls/APLAX/normalized/complete/2	22	83	14	77
gls-rls/APLAX/normalized/complete/2	23	82	15	76
gls-rls/all-views/z-normalized/ward/2	51	59	38	48
rls/APLAX/normalized/ward/2	24	81	16	75
gls/all-views/z-normalized/ward/2	54	55	42	45
gls/4CH/z-normalized/complete/2	47	61	36	52
gls-rls/all-views/z-normalized/complete/2	49	59	38	50
rls/2CH/normalized/ward/2	74	32	65	25
gls/4CH/normalized/ward/2	39	67	30	60
rls/4CH/normalized/complete/2	77	28	69	22
rls/2CH/normalized/complete/2	77	28	69	22
gls/APLAX/z-normalized/complete/2	40	65	32	59
gls-rls/APLAX/z-normalized/complete/2	42	63	34	57
gls-rls/APLAX/z-normalized/ward/2	43	62	35	56
rls/all-views/z-normalized/complete/2	48	57	40	51

gls/all-views/normalized/ward/2	37	67	30	62
gls-rls/2CH/normalized/ward/2	65	39	58	34
rls/2CH/z-normalized/ward/2	49	55	42	50
gls/APLAX/z-normalized/ward/2	37	66	31	62
gls-rls/all-views/normalized/complete/2	15	85	12	84
gls-rls/2CH/normalized/complete/2	70	33	64	29
rls/4CH/normalized/ward/2	79	23	74	20
gls/2CH/normalized/ward/2	62	41	56	37
rls/APLAX/normalized/complete/2	34	68	29	65
gls-rls/4CH/normalized/complete/2	35	67	30	64
rls/APLAX/z-normalized/complete/2	60	42	55	39
rls/APLAX/z-normalized/ward/2	30	70	27	69
gls/4CH/z-normalized/ward/2	33	67	30	66
gls-rls/APLAX/normalized/weighted/2	78	22	75	21
gls/APLAX/normalized/ward/2	76	24	73	23
gls/all-views/z-normalized/complete/2	64	35	62	35
gls/all-views/normalized/complete/2	84	15	82	15
rls/all-views/regular/median/2	94	5	92	5
gls-rls/2CH/z-normalized/weighted/2	97	2	95	2
rls/4CH/scaled/median/2	98	1	96	1
rls/4CH/regular/average/2	98	1	96	1
gls-rls/4CH/regular/single/2	98	0	97	1
gls-rls/all-views/scaled/median/2	98	0	97	1
gls-rls/all-views/z-normalized/centroid/2	98	0	97	1
gls-rls/APLAX/z-normalized/weighted/2	98	0	97	1
gls/all-views/normalized/single/2	98	0	97	1
gls-rls/APLAX/normalized/centroid/2	98	0	97	1
gls-rls/APLAX/normalized/average/2	98	0	97	1
gls-rls/all-views/scaled/single/2	98	0	97	1
gls-rls/APLAX/z-normalized/single/2	98	0	97	1
gls-rls/APLAX/normalized/median/2	98	0	97	1
gls-rls/APLAX/normalized/single/2	98	0	97	1
gls/all-views/z-normalized/single/2	98	0	97	1
gls-rls/APLAX/z-normalized/centroid/2	98	0	97	1
gls-rls/2CH/normalized/centroid/2	98	0	97	1
gls-rls/4CH/normalized/single/2	98	0	97	1
gls-rls/2CH/z-normalized/centroid/2	98	0	97	1
gls-rls/2CH/normalized/average/2	98	0	97	1
gls-rls/2CH/normalized/median/2	98	0	97	1
gls-rls/2CH/z-normalized/single/2	98	0	97	1
gls-rls/2CH/normalized/single/2	98	0	97	1
gls-rls/2CH/z-normalized/average/2	98	0	97	1
gls-rls/2CH/regular/weighted/2	98	0	97	1
gls-rls/2CH/regular/median/2	98	0	97	1
gls-rls/2CH/regular/centroid/2	98	0	97	1
gls/all-views/normalized/centroid/2	98	0	97	1
gls-rls/2CH/regular/average/2	98	0	97	1
gls/all-views/normalized/median/2	98	0	97	1
gls-rls/2CH/regular/single/2	98	0	97	1
gls/all-views/normalized/weighted/2	98	0	97	1
gls-rls/2CH/scaled/single/2	98	0	97	1

gls-rls/4CH/normalized/average/2	98	0	97	1
gls-rls/4CH/scaled/single/2	98	0	97	1
gls-rls/4CH/z-normalized/weighted/2	98	0	97	1
gls-rls/4CH/z-normalized/median/2	98	0	97	1
gls-rls/4CH/z-normalized/centroid/2	98	0	97	1
gls-rls/2CH/scaled/average/2	98	0	97	1
gls/all-views/normalized/average/2	98	0	97	1
gls-rls/4CH/z-normalized/average/2	98	0	97	1
gls-rls/4CH/z-normalized/complete/2	98	0	97	1
gls-rls/4CH/z-normalized/single/2	98	0	97	1
gls-rls/2CH/scaled/centroid/2	98	0	97	1
gls-rls/4CH/normalized/median/2	98	0	97	1
gls-rls/4CH/normalized/centroid/2	98	0	97	1
gls-rls/2CH/scaled/weighted/2	98	0	97	1
gls-rls/4CH/normalized/weighted/2	98	0	97	1
gls/4CH/normalized/median/2	98	0	97	1
gls-rls/all-views/z-normalized/single/2	98	0	97	1
gls/2CH/z-normalized/median/2	98	0	97	1
gls/APLAX/normalized/single/2	98	0	97	1
gls/APLAX/normalized/average/2	98	0	97	1
gls/APLAX/normalized/centroid/2	98	0	97	1
gls/APLAX/normalized/median/2	98	0	97	1
gls/APLAX/normalized/weighted/2	98	0	97	1
gls/APLAX/z-normalized/single/2	98	0	97	1
gls/APLAX/z-normalized/average/2	98	0	97	1
gls/APLAX/z-normalized/centroid/2	98	0	97	1
rls/all-views/regular/single/2	98	0	97	1
rls/all-views/regular/weighted/2	98	0	97	1
rls/all-views/normalized/single/2	98	0	97	1
rls/all-views/normalized/centroid/2	98	0	97	1
rls/all-views/normalized/median/2	98	0	97	1
rls/all-views/z-normalized/single/2	98	0	97	1
rls/all-views/z-normalized/centroid/2	98	0	97	1
rls/all-views/z-normalized/median/2	98	0	97	1
rls/all-views/scaled/single/2	98	0	97	1
gls/2CH/z-normalized/weighted/2	98	0	97	1
gls/2CH/z-normalized/centroid/2	98	0	97	1
rls/4CH/regular/median/2	98	0	97	1
gls/2CH/z-normalized/average/2	98	0	97	1
gls/4CH/z-normalized/single/2	98	0	97	1
gls/4CH/z-normalized/average/2	98	0	97	1
gls/4CH/z-normalized/centroid/2	98	0	97	1
gls/4CH/z-normalized/median/2	98	0	97	1
gls/4CH/z-normalized/weighted/2	98	0	97	1
gls/4CH/normalized/centroid/2	98	0	97	1
gls/4CH/normalized/average/2	98	0	97	1
gls/4CH/normalized/complete/2	98	0	97	1
gls/4CH/normalized/single/2	98	0	97	1
gls/2CH/normalized/single/2	98	0	97	1
gls/2CH/normalized/complete/2	98	0	97	1
gls/2CH/normalized/average/2	98	0	97	1

gls/2CH/normalized/centroid/2	98	0	97	1
gls/2CH/normalized/median/2	98	0	97	1
gls/2CH/normalized/weighted/2	98	0	97	1
gls/2CH/z-normalized/single/2	98	0	97	1
gls/2CH/z-normalized/complete/2	98	0	97	1
rls/4CH/regular/single/2	98	0	97	1
rls/4CH/normalized/single/2	98	0	97	1
gls-rls/all-views/normalized/centroid/2	98	0	97	1
rls/2CH/z-normalized/single/2	98	0	97	1
gls/4CH/normalized/weighted/2	98	0	97	1
rls/2CH/scaled/single/2	98	0	97	1
rls/2CH/scaled/average/2	98	0	97	1
gls/all-views/z-normalized/centroid/2	98	0	97	1
rls/2CH/scaled/centroid/2	98	0	97	1
rls/2CH/scaled/median/2	98	0	97	1
rls/2CH/scaled/weighted/2	98	0	97	1
rls/APLAX/normalized/single/2	98	0	97	1
rls/APLAX/normalized/centroid/2	98	0	97	1
rls/APLAX/normalized/weighted/2	98	0	97	1
rls/APLAX/z-normalized/single/2	98	0	97	1
rls/APLAX/z-normalized/centroid/2	98	0	97	1
rls/APLAX/z-normalized/median/2	98	0	97	1
gls/all-views/z-normalized/average/2	98	0	97	1
gls-rls/all-views/regular/single/2	98	0	97	1
gls-rls/all-views/regular/median/2	98	0	97	1
gls-rls/all-views/normalized/single/2	98	0	97	1
rls/2CH/z-normalized/average/2	98	0	97	1
rls/2CH/normalized/centroid/2	98	0	97	1
rls/4CH/normalized/average/2	98	0	97	1
rls/2CH/normalized/average/2	98	0	97	1
rls/4CH/normalized/centroid/2	98	0	97	1
rls/4CH/normalized/median/2	98	0	97	1
rls/4CH/normalized/weighted/2	98	0	97	1
rls/4CH/z-normalized/single/2	98	0	97	1
rls/4CH/z-normalized/complete/2	98	0	97	1
rls/4CH/z-normalized/average/2	98	0	97	1
rls/4CH/z-normalized/centroid/2	98	0	97	1
rls/4CH/z-normalized/median/2	98	0	97	1
rls/4CH/z-normalized/weighted/2	98	0	97	1
rls/4CH/scaled/single/2	98	0	97	1
gls/all-views/z-normalized/weighted/2	98	0	97	1
rls/2CH/regular/single/2	98	0	97	1
gls/all-views/z-normalized/median/2	98	0	97	1
rls/2CH/regular/average/2	98	0	97	1
rls/2CH/regular/centroid/2	98	0	97	1
rls/2CH/regular/weighted/2	98	0	97	1
rls/2CH/normalized/single/2	98	0	97	1
rls/2CH/z-normalized/centroid/2	98	0	97	1
gls/all-views/regular/single/2	99	1	96	0
gls/all-views/scaled/single/2	99	1	96	0
gls/4CH/regular/single/2	99	1	96	0

gls/4CH/regular/average/2	99	1	96	0
gls/4CH/scaled/single/2	99	1	96	0
gls/4CH/scaled/average/2	99	1	96	0
gls/2CH/regular/single/2	99	1	96	0
gls/2CH/scaled/single/2	99	1	96	0
gls/APLAX/regular/single/2	99	1	96	0
rls/all-views/regular/centroid/2	99	0	97	0
rls/all-views/normalized/average/2	99	1	96	0
rls/all-views/z-normalized/average/2	2	97	0	97
rls/all-views/z-normalized/weighted/2	2	97	0	97
rls/all-views/scaled/centroid/2	99	0	97	0
rls/4CH/regular/centroid/2	99	1	96	0
rls/4CH/scaled/centroid/2	99	1	96	0
rls/4CH/scaled/weighted/2	99	1	96	0
rls/2CH/regular/median/2	99	0	97	0
rls/2CH/normalized/weighted/2	99	1	96	0
rls/2CH/z-normalized/complete/2	3	97	0	96
rls/2CH/z-normalized/median/2	99	2	95	0
rls/APLAX/regular/single/2	99	1	96	0
rls/APLAX/regular/average/2	99	1	96	0
rls/APLAX/regular/centroid/2	99	0	97	0
rls/APLAX/regular/median/2	99	1	96	0
rls/APLAX/normalized/average/2	99	2	95	0
rls/APLAX/z-normalized/average/2	2	97	0	97
rls/APLAX/z-normalized/weighted/2	99	1	96	0
rls/APLAX/scaled/single/2	99	1	96	0
rls/APLAX/scaled/centroid/2	99	0	97	0
rls/APLAX/scaled/median/2	99	1	96	0
gls-rls/all-views/normalized/average/2	2	97	0	97
gls-rls/all-views/normalized/median/2	99	1	96	0
gls-rls/all-views/normalized/weighted/2	99	1	96	0
gls-rls/all-views/z-normalized/average/2	2	97	0	97
gls-rls/all-views/z-normalized/median/2	99	0	97	0
gls-rls/all-views/z-normalized/weighted/2	2	97	0	97
gls-rls/4CH/regular/centroid/2	99	1	96	0
gls-rls/4CH/regular/median/2	99	1	96	0
gls-rls/4CH/scaled/centroid/2	99	1	96	0
gls-rls/2CH/normalized/weighted/2	99	1	96	0
gls-rls/2CH/z-normalized/median/2	99	2	95	0
gls-rls/2CH/scaled/median/2	99	0	97	0
gls-rls/APLAX/regular/single/2	99	1	96	0
gls-rls/APLAX/z-normalized/average/2	2	97	0	97
gls-rls/APLAX/z-normalized/median/2	99	0	97	0
gls-rls/APLAX/scaled/single/2	99	1	96	0
gls-rls/APLAX/scaled/median/2	99	1	96	0

Table 10.2: Classification results of applying TSC to identify patient diagnoses. The results are sorted in descending order of DOR, although DOR is not included.

Dataset-Method	TP	TN	FP	FN
gls/2CH/regular/centroid/2	119	27	2	48

gls/2CH/scaled/centroid/2	119	27	2	48
gls/2CH/scaled/average/2	116	27	2	51
gls/2CH/regular/average/2	116	27	2	51
gls/2CH/scaled/ward/2	112	27	2	55
gls/2CH/regular/ward/2	112	27	2	55
gls-rls/2CH/scaled/ward/2	111	27	2	56
gls-rls/2CH/regular/ward/2	111	27	2	56
rls/all-views/normalized/weighted/2	166	4	25	1
rls/2CH/scaled/ward/2	106	27	2	61
rls/all-views/regular/complete/2	130	25	4	37
rls/4CH/regular/weighted/2	165	6	23	2
gls/all-views/regular/centroid/2	102	27	2	65
gls/2CH/scaled/complete/2	102	27	2	65
gls/2CH/regular/complete/2	102	27	2	65
gls/all-views/regular/weighted/2	136	24	5	31
gls/all-views/scaled/average/2	101	27	2	66
gls-rls/2CH/regular/complete/2	100	27	2	67
rls/APLAX/scaled/average/2	116	26	3	51
gls-rls/2CH/scaled/complete/2	99	27	2	68
gls-rls/4CH/scaled/weighted/2	130	24	5	37
rls/2CH/regular/complete/2	92	27	2	75
gls/all-views/scaled/ward/2	91	27	2	76
gls/all-views/regular/ward/2	91	27	2	76
gls/all-views/scaled/complete/2	90	27	2	77
gls/APLAX/regular/centroid/2	90	27	2	77
gls/4CH/scaled/centroid/2	107	26	3	60
gls/4CH/regular/centroid/2	107	26	3	60
gls/APLAX/regular/median/2	89	27	2	78
gls/APLAX/regular/ward/2	89	27	2	78
gls-rls/4CH/regular/complete/2	145	20	9	22
gls-rls/APLAX/scaled/average/2	117	25	4	50
gls/APLAX/regular/average/2	86	27	2	81
gls/4CH/regular/complete/2	104	26	3	63
gls/4CH/scaled/complete/2	104	26	3	63
gls-rls/4CH/scaled/median/2	164	6	23	3
gls-rls/all-views/regular/centroid/2	84	27	2	83
rls/2CH/scaled/complete/2	84	27	2	83
gls/all-views/scaled/centroid/2	83	27	2	84
gls-rls/all-views/scaled/centroid/2	83	27	2	84
gls/all-views/regular/complete/2	83	27	2	84
gls/all-views/regular/average/2	83	27	2	84
rls/APLAX/scaled/weighted/2	135	22	7	32
gls-rls/all-views/regular/ward/2	82	27	2	85
gls-rls/all-views/scaled/average/2	81	27	2	86
gls-rls/all-views/scaled/weighted/2	80	27	2	87
gls-rls/all-views/regular/complete/2	80	27	2	87
gls-rls/4CH/scaled/average/2	80	27	2	87
gls-rls/2CH/z-normalized/complete/2	166	2	27	1
rls/2CH/z-normalized/weighted/2	166	2	27	1
rls/APLAX/scaled/ward/2	79	27	2	88
rls/all-views/scaled/complete/2	79	27	2	88

rls/APLAX/scaled/complete/2	79	27	2	88
gls/2CH/scaled/weighted/2	78	27	2	89
gls-rls/all-views/regular/average/2	78	27	2	89
gls/2CH/scaled/median/2	78	27	2	89
gls-rls/APLAX/regular/average/2	78	27	2	89
rls/all-views/scaled/ward/2	78	27	2	89
gls/2CH/regular/median/2	78	27	2	89
gls/2CH/regular/weighted/2	78	27	2	89
rls/APLAX/regular/ward/2	78	27	2	89
rls/all-views/scaled/weighted/2	77	27	2	90
gls-rls/APLAX/scaled/centroid/2	77	27	2	90
gls-rls/APLAX/scaled/weighted/2	136	21	8	31
gls-rls/4CH/regular/weighted/2	164	5	24	3
rls/4CH/scaled/average/2	75	27	2	92
gls-rls/4CH/scaled/ward/2	75	27	2	92
rls/all-views/regular/ward/2	74	27	2	93
gls-rls/APLAX/regular/ward/2	74	27	2	93
gls-rls/all-views/scaled/ward/2	74	27	2	93
gls-rls/4CH/regular/ward/2	73	27	2	94
gls-rls/4CH/regular/average/2	73	27	2	94
rls/APLAX/regular/complete/2	73	27	2	94
gls-rls/all-views/scaled/complete/2	72	27	2	95
rls/4CH/regular/ward/2	72	27	2	95
rls/all-views/regular/average/2	72	27	2	95
gls-rls/APLAX/scaled/ward/2	72	27	2	95
rls/all-views/scaled/average/2	72	27	2	95
gls-rls/4CH/scaled/complete/2	72	27	2	95
rls/4CH/regular/complete/2	89	26	3	78
gls-rls/APLAX/regular/complete/2	107	24	5	60
rls/4CH/scaled/complete/2	81	26	3	86
gls/all-views/scaled/median/2	93	25	4	74
gls-rls/2CH/z-normalized/weighted/2	165	2	27	2
rls/4CH/regular/average/2	166	1	28	1
rls/4CH/scaled/median/2	166	1	28	1
gls/APLAX/normalized/ward/2	134	14	15	33
rls/2CH/normalized/ward/2	126	16	13	41
rls/4CH/normalized/ward/2	137	13	16	30
rls/2CH/normalized/complete/2	131	14	15	36
gls-rls/APLAX/normalized/weighted/2	136	12	17	31
rls/4CH/normalized/complete/2	130	13	16	37
gls-rls/2CH/normalized/ward/2	111	17	12	56
gls-rls/2CH/normalized/complete/2	120	15	14	47
rls/APLAX/z-normalized/ward/2	124	14	15	43
gls/all-views/z-normalized/complete/2	113	16	13	54
gls-rls/4CH/normalized/complete/2	116	14	15	51
gls/all-views/normalized/ward/2	114	14	15	53
gls/all-views/normalized/complete/2	144	7	22	23
gls/APLAX/z-normalized/ward/2	113	14	15	54
gls/4CH/z-normalized/ward/2	117	13	16	50
gls/APLAX/z-normalized/complete/2	109	14	15	58
gls/4CH/normalized/ward/2	111	13	16	56

rls/2CH/z-normalized/ward/2	92	16	13	75
gls-rls/APLAX/z-normalized/ward/2	103	14	15	64
gls-rls/all-views/z-normalized/ward/2	93	15	14	74
gls-rls/2CH/z-normalized/ward/2	120	10	19	47
gls/all-views/z-normalized/ward/2	87	16	13	80
gls/4CH/z-normalized/complete/2	98	14	15	69
rls/all-views/z-normalized/ward/2	95	14	15	72
rls/APLAX/normalized/complete/2	114	10	19	53
gls-rls/APLAX/normalized/complete/2	135	6	23	32
gls-rls/4CH/normalized/ward/2	95	13	16	72
gls-rls/4CH/z-normalized/ward/2	83	15	14	84
rls/all-views/normalized/ward/2	83	15	14	84
rls/all-views/z-normalized/complete/2	92	13	16	75
rls/4CH/z-normalized/ward/2	86	14	15	81
gls-rls/APLAX/normalized/ward/2	132	6	23	35
gls/APLAX/normalized/complete/2	136	5	24	31
rls/APLAX/z-normalized/complete/2	97	11	18	70
gls/all-views/scaled/weighted/2	153	2	27	14
rls/APLAX/normalized/ward/2	132	5	24	35
gls/2CH/z-normalized/ward/2	105	9	20	62
gls-rls/APLAX/z-normalized/complete/2	100	9	20	67
rls/all-views/regular/median/2	158	1	28	9
gls-rls/all-views/z-normalized/complete/2	90	10	19	77
rls/all-views/normalized/complete/2	120	5	24	47
gls/2CH/normalized/ward/2	97	8	21	70
gls/all-views/regular/median/2	144	2	27	23
gls-rls/all-views/normalized/complete/2	142	2	27	25
gls/4CH/scaled/median/2	139	2	27	28
gls/4CH/scaled/weighted/2	139	2	27	28
gls/4CH/regular/weighted/2	139	2	27	28
gls/4CH/regular/median/2	139	2	27	28
gls/APLAX/regular/weighted/2	122	2	27	45
gls-rls/APLAX/regular/median/2	138	1	28	29
gls-rls/APLAX/scaled/complete/2	116	2	27	51
gls/4CH/scaled/ward/2	111	2	27	56
gls/4CH/regular/ward/2	111	2	27	56
rls/APLAX/regular/weighted/2	108	2	27	59
gls/APLAX/regular/complete/2	107	2	27	60
gls-rls/all-views/regular/weighted/2	106	2	27	61
rls/2CH/regular/ward/2	106	2	27	61
gls-rls/APLAX/regular/weighted/2	128	1	28	39
gls-rls/APLAX/regular/centroid/2	99	2	27	68
rls/4CH/scaled/ward/2	97	2	27	70
gls/4CH/z-normalized/weighted/2	166	0	29	1
gls-rls/4CH/scaled/single/2	166	0	29	1
gls/2CH/normalized/median/2	166	0	29	1
gls/2CH/normalized/centroid/2	166	0	29	1
gls/2CH/normalized/average/2	166	0	29	1
gls/2CH/normalized/complete/2	166	0	29	1
gls/2CH/normalized/single/2	166	0	29	1
gls-rls/2CH/regular/single/2	166	0	29	1

rls/4CH/scaled/single/2	166	0	29	1
gls-rls/4CH/z-normalized/weighted/2	166	0	29	1
gls/4CH/z-normalized/median/2	166	0	29	1
gls-rls/2CH/regular/centroid/2	166	0	29	1
gls-rls/2CH/regular/median/2	166	0	29	1
gls-rls/2CH/regular/weighted/2	166	0	29	1
gls-rls/2CH/normalized/single/2	166	0	29	1
gls/4CH/z-normalized/centroid/2	166	0	29	1
gls-rls/2CH/normalized/average/2	166	0	29	1
gls-rls/2CH/regular/average/2	166	0	29	1
gls-rls/4CH/z-normalized/median/2	166	0	29	1
gls-rls/2CH/normalized/centroid/2	166	0	29	1
gls-rls/4CH/normalized/average/2	166	0	29	1
gls-rls/4CH/regular/single/2	166	0	29	1
gls/2CH/z-normalized/weighted/2	166	0	29	1
gls/2CH/z-normalized/median/2	166	0	29	1
gls/2CH/z-normalized/centroid/2	166	0	29	1
gls/2CH/z-normalized/average/2	166	0	29	1
gls-rls/4CH/normalized/single/2	166	0	29	1
gls/2CH/z-normalized/complete/2	166	0	29	1
gls/2CH/z-normalized/single/2	166	0	29	1
gls-rls/4CH/z-normalized/centroid/2	166	0	29	1
gls-rls/4CH/normalized/centroid/2	166	0	29	1
gls-rls/4CH/normalized/median/2	166	0	29	1
gls-rls/4CH/normalized/weighted/2	166	0	29	1
gls-rls/4CH/z-normalized/single/2	166	0	29	1
gls-rls/4CH/z-normalized/complete/2	166	0	29	1
gls-rls/4CH/z-normalized/average/2	166	0	29	1
gls/2CH/normalized/weighted/2	166	0	29	1
gls/4CH/z-normalized/average/2	166	0	29	1
gls-rls/2CH/z-normalized/single/2	166	0	29	1
gls-rls/2CH/normalized/median/2	166	0	29	1
gls/all-views/normalized/weighted/2	166	0	29	1
gls/all-views/z-normalized/centroid/2	166	0	29	1
gls-rls/APLAX/normalized/centroid/2	166	0	29	1
gls-rls/APLAX/normalized/median/2	166	0	29	1
gls/all-views/z-normalized/average/2	166	0	29	1
gls-rls/APLAX/z-normalized/single/2	166	0	29	1
gls/all-views/z-normalized/single/2	166	0	29	1
gls-rls/APLAX/z-normalized/average/2	165	0	29	2
gls-rls/APLAX/z-normalized/centroid/2	166	0	29	1
gls-rls/all-views/scaled/median/2	166	0	29	1
gls-rls/APLAX/z-normalized/weighted/2	166	0	29	1
gls-rls/APLAX/scaled/single/2	166	0	29	1
gls/all-views/normalized/median/2	166	0	29	1
gls/all-views/normalized/centroid/2	166	0	29	1
gls/all-views/normalized/average/2	166	0	29	1
gls/all-views/normalized/single/2	166	0	29	1
gls-rls/APLAX/scaled/median/2	166	0	29	1
gls-rls/APLAX/normalized/average/2	166	0	29	1
gls/all-views/z-normalized/median/2	166	0	29	1

gls-rls/APLAX/normalized/single/2	166	0	29	1
gls/all-views/z-normalized/weighted/2	166	0	29	1
gls/4CH/z-normalized/single/2	166	0	29	1
gls-rls/2CH/z-normalized/average/2	166	0	29	1
gls/4CH/normalized/weighted/2	166	0	29	1
gls-rls/2CH/z-normalized/centroid/2	166	0	29	1
gls/4CH/normalized/median/2	166	0	29	1
gls-rls/2CH/scaled/single/2	166	0	29	1
gls/4CH/normalized/centroid/2	166	0	29	1
gls-rls/2CH/scaled/average/2	166	0	29	1
gls/4CH/normalized/average/2	166	0	29	1
gls-rls/2CH/scaled/centroid/2	166	0	29	1
gls-rls/2CH/scaled/weighted/2	166	0	29	1
gls-rls/APLAX/regular/single/2	166	0	29	1
gls/4CH/normalized/complete/2	166	0	29	1
gls/4CH/normalized/single/2	166	0	29	1
gls/all-views/scaled/single/2	166	0	29	1
gls/APLAX/regular/single/2	166	0	29	1
gls/APLAX/normalized/single/2	166	0	29	1
rls/4CH/z-normalized/weighted/2	166	0	29	1
rls/APLAX/regular/single/2	166	0	29	1
rls/2CH/scaled/single/2	166	0	29	1
rls/4CH/normalized/average/2	166	0	29	1
rls/2CH/scaled/average/2	166	0	29	1
rls/4CH/normalized/single/2	166	0	29	1
rls/2CH/scaled/centroid/2	166	0	29	1
rls/2CH/scaled/median/2	166	0	29	1
rls/2CH/scaled/weighted/2	166	0	29	1
rls/4CH/regular/median/2	166	0	29	1
rls/2CH/z-normalized/centroid/2	166	0	29	1
rls/APLAX/regular/average/2	166	0	29	1
rls/4CH/regular/single/2	166	0	29	1
rls/APLAX/regular/median/2	166	0	29	1
rls/all-views/scaled/median/2	164	0	29	3
rls/APLAX/normalized/single/2	166	0	29	1
rls/all-views/scaled/single/2	166	0	29	1
rls/APLAX/normalized/average/2	165	0	29	2
rls/4CH/normalized/centroid/2	166	0	29	1
rls/4CH/normalized/median/2	166	0	29	1
rls/APLAX/normalized/centroid/2	166	0	29	1
rls/2CH/regular/weighted/2	166	0	29	1
rls/4CH/z-normalized/median/2	166	0	29	1
rls/4CH/z-normalized/centroid/2	166	0	29	1
rls/2CH/regular/single/2	166	0	29	1
rls/4CH/z-normalized/average/2	166	0	29	1
rls/2CH/regular/average/2	166	0	29	1
rls/4CH/z-normalized/complete/2	166	0	29	1
rls/2CH/regular/centroid/2	166	0	29	1
rls/2CH/normalized/single/2	166	0	29	1
rls/2CH/z-normalized/average/2	166	0	29	1
rls/4CH/z-normalized/single/2	166	0	29	1

rls/2CH/normalized/average/2	166	0	29	1
rls/4CH/normalized/weighted/2	166	0	29	1
rls/2CH/normalized/centroid/2	166	0	29	1
rls/2CH/normalized/median/2	128	0	29	39
rls/2CH/z-normalized/single/2	166	0	29	1
rls/2CH/z-normalized/complete/2	164	0	29	3
rls/all-views/z-normalized/weighted/2	165	0	29	2
rls/APLAX/normalized/median/2	162	0	29	5
gls/APLAX/normalized/average/2	166	0	29	1
gls-rls/all-views/z-normalized/single/2	166	0	29	1
gls-rls/all-views/normalized/single/2	166	0	29	1
gls/APLAX/z-normalized/average/2	166	0	29	1
gls-rls/all-views/normalized/average/2	165	0	29	2
gls-rls/all-views/normalized/ward/2	128	0	29	39
gls-rls/all-views/normalized/centroid/2	166	0	29	1
gls-rls/all-views/normalized/median/2	166	0	29	1
gls-rls/all-views/normalized/weighted/2	166	0	29	1
gls/APLAX/z-normalized/single/2	166	0	29	1
gls-rls/all-views/regular/median/2	166	0	29	1
gls-rls/all-views/z-normalized/average/2	165	0	29	2
gls/APLAX/normalized/weighted/2	166	0	29	1
gls-rls/all-views/z-normalized/centroid/2	166	0	29	1
gls-rls/all-views/z-normalized/weighted/2	165	0	29	2
gls-rls/all-views/scaled/single/2	166	0	29	1
gls/APLAX/normalized/median/2	166	0	29	1
gls/APLAX/normalized/centroid/2	166	0	29	1
gls/APLAX/z-normalized/centroid/2	166	0	29	1
rls/all-views/regular/single/2	166	0	29	1
rls/APLAX/normalized/weighted/2	166	0	29	1
rls/APLAX/scaled/single/2	166	0	29	1
rls/APLAX/z-normalized/single/2	166	0	29	1
rls/all-views/z-normalized/median/2	166	0	29	1
rls/APLAX/z-normalized/average/2	165	0	29	2
rls/all-views/z-normalized/centroid/2	166	0	29	1
rls/APLAX/z-normalized/centroid/2	166	0	29	1
rls/APLAX/z-normalized/median/2	166	0	29	1
rls/APLAX/z-normalized/weighted/2	166	0	29	1
rls/all-views/z-normalized/average/2	165	0	29	2
rls/all-views/regular/weighted/2	166	0	29	1
rls/all-views/z-normalized/single/2	166	0	29	1
rls/all-views/normalized/median/2	166	0	29	1
rls/APLAX/scaled/median/2	166	0	29	1
rls/all-views/normalized/centroid/2	166	0	29	1
gls-rls/all-views/regular/single/2	166	0	29	1
rls/all-views/normalized/average/2	166	0	29	1
rls/all-views/normalized/single/2	166	0	29	1
gls/all-views/regular/single/2	166	0	29	1
gls/4CH/regular/single/2	167	1	28	0
gls/4CH/regular/average/2	167	1	28	0
gls/4CH/scaled/single/2	167	1	28	0
gls/4CH/scaled/average/2	167	1	28	0

gls/2CH/regular/single/2	167	1	28	0
gls/2CH/scaled/single/2	167	1	28	0
rls/all-views/regular/centroid/2	167	0	29	0
rls/all-views/scaled/centroid/2	167	0	29	0
rls/4CH/regular/centroid/2	167	1	28	0
rls/4CH/scaled/centroid/2	167	1	28	0
rls/4CH/scaled/weighted/2	167	1	28	0
rls/2CH/regular/median/2	167	0	29	0
rls/2CH/normalized/weighted/2	167	1	28	0
rls/2CH/z-normalized/median/2	167	2	27	0
rls/APLAX/regular/centroid/2	167	0	29	0
rls/APLAX/scaled/centroid/2	167	0	29	0
gls-rls/all-views/z-normalized/median/2	167	0	29	0
gls-rls/4CH/regular/centroid/2	167	1	28	0
gls-rls/4CH/regular/median/2	167	1	28	0
gls-rls/4CH/scaled/centroid/2	167	1	28	0
gls-rls/2CH/normalized/weighted/2	167	1	28	0
gls-rls/2CH/z-normalized/median/2	167	2	27	0
gls-rls/2CH/scaled/median/2	167	0	29	0
gls-rls/APLAX/z-normalized/median/2	167	0	29	0

Table 10.3: Classification results of applying TSC to identify heart failure among patients. The results are sorted in descending order of DOR, although DOR is not included.

Preprocessing-Method	TP	TN	FP	FN
regular/weighted/2	822	1610	85	996
scaled/weighted/2	822	1610	85	996
regular/ward/2	1202	1491	204	616
scaled/ward/2	1202	1491	204	616
regular/complete/2	1133	1515	180	685
scaled/complete/2	1133	1515	180	685
z-norm/complete/2	471	1604	91	1347
z-norm/weighted/2	583	1553	142	1235
norm/ward/2	903	1049	646	915
z-norm/ward/2	1091	845	850	727
norm/complete/2	1704	58	1637	114
norm/weighted/2	1756	4	1691	62
regular/average/2	1816	0	1695	2
scaled/average/2	1816	0	1695	2
regular/centroid/2	1815	0	1695	3
scaled/centroid/2	1815	0	1695	3
z-norm/average/2	1814	0	1695	4
z-norm/centroid/2	1814	0	1695	4
norm/average/2	1809	0	1695	9
norm/centroid/2	1818	1	1694	0

10.1.2 Peak-value Clustering

Table 10.4: Classification results of applying PVC to identify heart failure among patients. The results are sorted in descending order of DOR, although DOR is not included.

Dataset-Method	TP	TN	FP	FN
gls-EF/ward/2	83	63	37	12
gls-EF/complete/2	77	72	28	18
gls-EF/average/2	81	65	35	14
rls-EF/complete/2	83	55	36	14
gls-rls-EF/ward/2	78	55	36	15
gls-rls-EF/complete/2	70	62	29	23
rls-EF/ward/2	58	74	17	39
rls/average/2	61	72	19	36
gls-rls/ward/2	56	71	20	37
rls/ward/2	57	71	20	40
gls/ward/2	59	74	26	36
gls-rls/complete/2	4	90	1	89
rls/complete/2	58	66	25	39
gls-rls/average/2	92	3	88	1
gls/complete/2	16	83	17	79
rls-EF/single/2	96	0	91	1
rls-EF/average/2	96	0	91	1
gls/average/2	0	99	1	95
gls/single/2	0	99	1	95
gls-rls-EF/single/2	92	0	91	1
gls-rls-EF/average/2	92	0	91	1
rls/single/2	97	1	90	0
gls-EF/single/2	1	100	0	94
gls-rls/single/2	93	1	90	0

Table 10.5: Classification results of applying PVC to identify patient diagnoses among patients. The results are sorted in descending order of DOR, although DOR is not included.

Dataset-Method	TP	TN	FP	FN
gls-EF/ward/2	118	30	2	45
rls-EF/complete/2	117	27	2	42
gls-rls-EF/ward/2	112	27	2	43
gls-EF/average/2	114	30	2	49
gls-EF/complete/2	103	30	2	60
gls-rls-EF/complete/2	97	27	2	58
rls/complete/2	81	27	2	78
rls/average/2	78	27	2	81
gls-rls/ward/2	74	27	2	81
rls/ward/2	75	27	2	84
rls-EF/ward/2	73	27	2	86
gls/ward/2	82	29	3	81
gls-rls/average/2	153	2	27	2
gls/complete/2	137	7	25	26
gls-rls/complete/2	150	0	29	5
gls-EF/single/2	162	0	32	1
rls-EF/single/2	158	0	29	1
rls/single/2	158	0	29	1
rls-EF/average/2	158	0	29	1
gls-rls-EF/single/2	154	0	29	1

gls-rls-EF/average/2	154	0	29	1
gls/single/2	163	1	31	0
gls/average/2	163	1	31	0
gls-rls/single/2	155	1	28	0

10.2 Neural Network

Table 10.6: Classification results of NN, when trained to predict heart failure among patients. The results are sorted in descending order of DOR, although DOR is not included.

Dataset-Model	TP	TN	FP	FN
gls/4CH/upsampled	46	61	39	53
rls/APLAX/regular	48	58	42	51
rls/4CH/regular	36	68	32	64
gls/APLAX/downsampled	62	40	60	36
gls/2CH/downsampled	60	39	58	39
gls/4CH/downsampled	48	52	48	51
gls/APLAX/regular	48	50	50	51
gls/2CH/regular	57	39	58	43
gls/4CH/regular	61	34	66	39
all-strain/4CH/regular	52	31	69	48
rls/APLAX/downsampled	33	47	53	65
all-strain/all-views/regular	53	27	70	46
rls/2CH/downsampled	30	45	52	69
all-strain/all-views/downsampled	36	36	61	62
gls/APLAX/upsampled	49	24	76	49
rls/2CH/regular	36	34	63	64
gls/2CH/upsampled	58	16	81	41
all-strain/4CH/upsampled	19	54	46	80
all-strain/2CH/downsampled	64	10	87	35
all-strain/APLAX/regular	41	22	78	58
all-strain/all-views/upsampled	25	33	64	73
all-strain/APLAX/downsampled	34	22	78	64
gls/all-views/regular	25	28	69	74
all-strain/2CH/upsampled	51	9	88	48
rls/all-views/downsampled	51	8	89	47
all-strain/4CH/downsampled	35	15	85	64
rls/4CH/upsampled	22	24	76	77
rls/4CH/downsampled	36	13	87	63
rls/APLAX/upsampled	27	16	84	71
rls/all-views/upsampled	13	29	68	85
gls/all-views/upsampled	13	29	68	85
gls/all-views/downsampled	46	6	91	52
rls/all-views/regular	27	13	84	72
rls/2CH/upsampled	32	9	88	67
all-strain/APLAX/upsampled	41	3	97	57
all-strain/2CH/regular	42	0	97	58

Table 10.7: Classification results of NN, when trained to predict patient diagnoses. The results are sorted in descending order of DOR, although DOR is not included.

Dataset-Preprocessing	TP	TN	FP	FN
all-strain/4CH/upsampled	166	0	32	1
all-strain/2CH/regular	168	0	29	0
gls/2CH/regular	168	0	29	0
rls/2CH/regular	168	0	29	0
all-strain/2CH/downsampled	167	0	29	0
all-strain/2CH/upsampled	167	0	29	0
gls/2CH/downsampled	167	0	29	0
gls/2CH/upsampled	167	0	29	0
rls/2CH/downsampled	167	0	29	0
rls/2CH/upsampled	167	0	29	0
all-strain/all-views/regular	167	0	29	0
gls/all-views/regular	167	0	29	0
rls/all-views/regular	167	0	29	0
all-strain/all-views/downsampled	166	0	29	0
all-strain/all-views/upsampled	166	0	29	0
gls/all-views/downsampled	166	0	29	0
gls/all-views/upsampled	166	0	29	0
rls/all-views/downsampled	166	0	29	0
rls/all-views/upsampled	166	0	29	0
all-strain/4CH/regular	168	0	32	0
gls/4CH/regular	168	0	32	0
rls/4CH/regular	168	0	32	0
all-strain/4CH/downsampled	167	0	32	0
gls/4CH/downsampled	167	0	32	0
gls/4CH/upsampled	167	0	32	0
rls/4CH/downsampled	167	0	32	0
rls/4CH/upsampled	167	0	32	0
all-strain/APLAX/regular	167	0	32	0
gls/APLAX/regular	167	0	32	0
rls/APLAX/regular	167	0	32	0
all-strain/APLAX/downsampled	166	0	32	0
all-strain/APLAX/upsampled	166	0	32	0
gls/APLAX/downsampled	166	0	32	0
gls/APLAX/upsampled	166	0	32	0
rls/APLAX/downsampled	166	0	32	0
rls/APLAX/upsampled	166	0	32	0

Table 10.8: Classification results of NN, when trained to predict segment indication. The results are sorted in descending order of DOR, although DOR is not included.

Preprocessing	TP	TN	FP	FN
regular	1364	1274	607	331
downsampled	1255	1390	473	440
upsampled	934	1365	498	761

10.2.1 Peak-value Supervised Classifiers

Table 10.9: Classification results of PVSC, when trained to predict heart failure among patients. The results are sorted in descending order of DOR, although DOR is not included.

Dataset-Model	TP	TN	FP	FN
gls-EF/Gaussian-Process	74	72	27	21
rls-EF/MLP	74	67	23	23
rls-EF/Linear-SVM	73	67	23	24
gls-EF/Ada-Boost	73	72	27	22
gls-EF/Naive-Bayes	72	73	26	23
gls-EF/Linear-SVM	71	74	25	24
rls-EF/Decision-Tree	76	62	28	21
gls-EF/KNN	70	73	26	25
gls-EF/Random-Forest	74	68	31	21
rls-EF/Extra-Trees	77	60	30	20
gls-rls-EF/Naive-Bayes	71	63	27	22
rls-EF/Naive-Bayes	72	65	25	25
rls/Naive-Bayes	73	64	26	24
gls-rls-EF/Linear-SVM	68	66	24	25
gls-rls-EF/Extra-Trees	72	61	29	21
gls-rls/Decision-Tree	71	62	28	22
gls-rls/Naive-Bayes	70	63	27	23
gls-EF/Discriminant-Analysis	67	74	25	28
gls-EF/Extra-Trees	69	72	27	26
gls-rls-EF/Ada-Boost	69	64	26	24
rls/KNN	79	55	35	18
gls-rls-EF/Random-Forest	70	62	28	23
gls-rls/Extra-Trees	72	59	31	21
gls/Gaussian-Process	70	69	30	25
rls/Ada-Boost	74	60	30	23
gls-rls/Ada-Boost	70	61	29	23
gls/Linear-SVM	70	68	31	25
rls/Linear-SVM	73	60	30	24
gls-EF/Decision-Tree	67	71	28	28
rls-EF/KNN	75	57	33	22
gls/Ada-Boost	70	67	32	25
gls/Naive-Bayes	69	68	31	26
gls-rls/Linear-SVM	67	62	28	26
rls/Extra-Trees	74	57	33	23
gls-rls-EF/KNN	69	59	31	24
rls-EF/Ada-Boost	68	63	27	29
rls-EF/Random-Forest	71	60	30	26
rls/Decision-Tree	74	56	34	23
gls-rls-EF/Decision-Tree	66	61	29	27
gls-rls/KNN	71	55	35	22
gls/Discriminant-Analysis	65	69	30	30
gls-rls/Random-Forest	67	59	31	26
gls-rls/MLP	65	61	29	28
gls/KNN	60	73	26	35
rls/MLP	64	64	26	33
rls/Random-Forest	69	58	32	28
rls-EF/Discriminant-Analysis	68	59	31	29

gls/Extra-Trees	64	67	32	31
rls/Discriminant-Analysis	67	59	31	30
gls-rls-EF/MLP	55	67	23	38
gls/Random-Forest	69	60	39	26
rls/Gaussian-Process	69	52	38	28
rls-EF/Gaussian-Process	69	52	38	28
gls-rls-EF/Discriminant-Analysis	57	61	29	36
gls-rls-EF/Gaussian-Process	64	54	36	29
gls/Decision-Tree	62	63	36	33
gls/RBF-SVM	43	76	23	52
gls-rls/Discriminant-Analysis	54	59	31	39
gls-EF/RBF-SVM	9	95	4	86
gls-EF/MLP	42	74	25	53
gls-rls/Gaussian-Process	59	49	41	34
gls/MLP	40	72	27	55
rls/RBF-SVM	97	0	90	0
gls-rls/RBF-SVM	93	0	90	0
rls-EF/RBF-SVM	97	0	90	0
gls-rls-EF/RBF-SVM	93	0	90	0

Table 10.10: Classification results of PVSC, when trained to predict patient diagnoses. The results are sorted in descending order of DOR, although DOR is not included.

Dataset-Model	TP	TN	FP	FN
gls-rls-EF/Ada-Boost	151	22	6	4
gls-rls/KNN	147	23	5	8
rls-EF/Extra-Trees	153	21	7	6
gls-rls-EF/Extra-Trees	150	20	8	5
gls-rls/Extra-Trees	150	20	8	5
gls-rls-EF/KNN	146	23	5	9
rls/Linear-SVM	155	18	10	4
rls-EF/Random-Forest	155	18	10	4
rls/Extra-Trees	154	19	9	5
gls-rls-EF/Linear-SVM	150	19	9	5
rls-EF/Gaussian-Process	150	22	6	9
rls-EF/Linear-SVM	154	18	10	5
rls-EF/KNN	149	22	6	10
rls-EF/Ada-Boost	153	19	9	6
gls-rls-EF/Gaussian-Process	144	22	6	11
rls/KNN	151	20	8	8
gls-rls/Decision-Tree	147	20	8	8
gls-rls/Linear-SVM	149	18	10	6
gls-rls/Random-Forest	148	18	10	7
rls/Random-Forest	154	15	13	5
rls/Ada-Boost	151	18	10	8
rls/Gaussian-Process	147	20	8	12
gls-rls-EF/Decision-Tree	143	20	8	12
gls-rls/Ada-Boost	149	15	13	6
rls/Naive-Bayes	121	25	3	38
gls-rls/Naive-Bayes	117	25	3	38
rls-EF/Naive-Bayes	120	25	3	39

gls-rls-EF/Naive-Bayes	116	25	3	39
gls-EF/Extra-Trees	154	18	13	9
gls-EF/Naive-Bayes	132	26	5	31
gls/Naive-Bayes	137	25	6	26
rls-EF/Decision-Tree	151	15	13	8
gls-rls-EF/Random-Forest	147	15	13	8
gls-rls/Gaussian-Process	142	18	10	13
gls-rls/MLP	145	16	12	10
rls/Decision-Tree	149	15	13	10
gls-EF/Random-Forest	152	16	15	11
gls-EF/KNN	148	18	13	15
rls-EF/MLP	151	11	17	8
gls/Extra-Trees	152	14	17	11
gls-EF/Gaussian-Process	162	2	29	1
gls-EF/Decision-Tree	147	17	14	16
gls/Random-Forest	153	13	18	10
gls/KNN	152	13	18	11
rls/Discriminant-Analysis	157	3	25	2
rls-EF/Discriminant-Analysis	157	3	25	2
gls-rls-EF/MLP	146	10	18	9
rls/MLP	148	11	17	11
gls/MLP	160	4	27	3
gls-EF/Ada-Boost	147	14	17	16
gls/Decision-Tree	147	14	17	16
gls-EF/Discriminant-Analysis	153	10	21	10
gls/Discriminant-Analysis	153	10	21	10
gls/Ada-Boost	147	13	18	16
gls-EF/MLP	158	5	26	5
gls-EF/Linear-SVM	161	2	29	2
gls/Linear-SVM	161	2	29	2
gls/Gaussian-Process	160	2	29	3
gls/RBF-SVM	163	1	30	0
rls/RBF-SVM	159	0	28	0
gls-rls/RBF-SVM	155	0	28	0
gls-rls/Discriminant-Analysis	155	1	27	0
gls-EF/RBF-SVM	163	0	31	0
rls-EF/RBF-SVM	159	0	28	0
gls-rls-EF/RBF-SVM	155	0	28	0
gls-rls-EF/Discriminant-Analysis	155	1	27	0